



## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets <sup>5</sup> : C12N 15/52, C07K 15/04 C12P 19/42, C12N 9/00, 1/21 C12P 19/42 // C12R 1/38 C12R 1/41C12R 1/01		A1	(11) Numéro de publication internationale: WO 91/11518
			(43) Date de publication internationale: 8 août 1991 (08.08.91)
(21) Numéro de la demande internationale: PCT/FR91/00054		(74) Mandataire: BECKER, Philippe; Rhône-Poulenc Rorer S.A. - Direction Brevets, 20, avenue Raymond-Aron, F-92165 Antony Cédex (FR).	
(22) Date de dépôt international: 30 janvier 1991 (30.01.91)		(81) Etats désignés: AT (brevet européen), BE (brevet européen), CA, CH (brevet européen), DE (brevet européen), DK (brevet européen), ES (brevet européen), FR (brevet européen), GB (brevet européen), GR (brevet européen), HU, IT (brevet européen), JP, LU (brevet européen), NL (brevet européen), SE (brevet européen), SU, US.	
(30) Données relatives à la priorité: 90/01137 31 janvier 1990 (31.01.90) FR		(Publiée) <i>Avec rapport de recherche internationale. Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont reçues.</i>	
(71) Déposant ( <i>pour tous les Etats désignés sauf US</i> ): RHÔNE-POULENC BIOCHIMIE [FR/FR]; 20, avenue Raymond-Aron, F-92160 Antony (FR).			
(72) Inventeurs; et			
(75) Inventeurs/Déposants ( <i>US seulement</i> ): BLANCHE, Francis [FR/FR]; 41, rue des Solitaires, F-75019 Paris (FR). CAMERON, Béatrice [FR/FR]; 28, rue Henri-Barbusse, F-75005 Paris (FR). CROUZET, Joël [FR/FR]; 48-52, rue des Meuniers, F-75012 Paris (FR). DEBUSSCHE, Laurent [FR/FR]; 43, bd de l'Hôpital, F-75013 Paris (FR). LEVY-SCHIL, Sophie [FR/FR]; 2, rue Monttessuy, F-75007 Paris (FR). THIBAUT, Denis [FR/FR]; 28, rue Jean-Colly, F-75013 Paris (FR).			
(54) Title: POLYPEPTIDES INVOLVED IN THE BIOSYNTHESIS OF COBALAMINES AND/OR COBAMIDES, DNA SEQUENCES CODING FOR THESE POLYPEPTIDES, AND THEIR PREPARATION AND USE			
(54) Titre: POLYPEPTIDES IMPLIQUES DANS LA BIOSYNTHESE DES COBALAMINES ET/OU DES COBAMIDES, SEQUENCES D'ADN CODANT POUR CES POLYPEPTIDES, PROCÉDÉ DE PRÉPARATION, ET LEUR UTILISATION			
(57) Abstract			
<p>Novel polypeptides involved in the biosynthesis of cobalamines and/or cobamides, in particular coenzyme B<sub>12</sub>, genetic material responsible for expressing these polypeptides, and a method for preparing them, are described. A method for amplifying the production of cobalamines, and particularly coenzyme B<sub>12</sub>, using recombinant DNA techniques, are also described.</p>			
(57) Abrégé			
<p>La présente invention concerne de nouveaux polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides, et en particulier du coenzyme B<sub>12</sub>. Elle concerne également le matériel génétique responsable de l'expression de ces polypeptides, ainsi qu'un procédé permettant leur préparation. Elle concerne enfin un procédé d'amplification de la production de cobalamines, et plus particulièrement du coenzyme B<sub>12</sub>, par les techniques d'ADN recombinant.</p>			
BEST AVAILABLE COPY			

***UNIQUEMENT A TITRE D'INFORMATION***

**Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publient des demandes internationales en vertu du PCT.**

AT	Autriche	ES	Espagne	MC	Madagascar
AU	Australie	FI	Finlande	ML	Mali
BB	Barbade	FR	France	MN	Mongolie
BE	Belgique	GA	Gabon	MR	Mauritanie
BF	Burkina Faso	GB	Royaume-Uni	MW	Malawi
BG	Bulgarie	GN	Guinée	NL	Pays-Bas
BJ	Bénin	GR	Grèce	NO	Norvège
BR	Brésil	HU	Hongrie	PL	Pologne
CA	Canada	IT	Italic	RO	Roumanie
CF	République Centrafricaine	JP	Japon	SD	Soudan
CG	Congo	KP	République populaire démocratique de Corée	SE	Suède
CH	Suisse	KR	République de Corée	SN	Sénégal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Union soviétique
CM	Cameroun	LK	Sri Lanka	TD	Tchad
CS	Tchécoslovaquie	LU	Luxembourg	TG	Togo
DE	Allemagne	MC	Monaco	US	Etats-Unis d'Amérique
DK	Danemark				

POLYPEPTIDES IMPLIQUES DANS LA BIOSYNTHESE DES COBALAMINES  
ET/OU DES COBAMIDES, SEQUENCES D'ADN CODANT POUR CES POLYPEPTIDES,  
PROCEDE DE PREPARATION, ET LEUR UTILISATION

La présente invention concerne de nouveaux polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides, et en particulier du coenzyme B<sub>12</sub>. Elle concerne également le matériel génétique responsable de l'expression de ces polypeptides, ainsi qu'un 5 procédé permettant leur préparation. Elle concerne enfin un procédé d'amplification de la production de cobalamines, et plus particulièrement du coenzyme B<sub>12</sub>, par les techniques d'ADN recombinant.

La vitamine B<sub>12</sub> fait partie des vitamines du groupe B. Il 10 s'agit d'une vitamine hydrosoluble qui a été identifiée comme étant le facteur permettant de traiter des malades souffrant d'anémie pernicieuse. Elle est généralement prescrite pour stimuler l'hématopoïèse chez les sujets fatigués, mais elle est aussi utilisée dans de nombreux autres cas qui comprennent les troubles hépatiques, 15 des déficiences nerveuses ou comme stimulant de l'appétit, principe tonifiant, ainsi qu'en dermatologie (Beck, 1982, Fraser et al., 1983). Dans les élevages industriels d'animaux non ruminants, l'alimentation étant essentiellement à base de protéines d'origines végétales, il est nécessaire d'incorporer aux rations alimentaires de la vitamine B<sub>12</sub>, à 20 des quantités de 10 à 15 mg par tonne d'aliments (Barrère et al., 1981).

La vitamine B<sub>12</sub> fait partie d'une classe de molécules appelées cobalamines, dont la structure est présentée à la figure 1. Les cobamides diffèrent des cobalamines par la base du nucléotide 25 inférieur qui n'est plus le 5,6-diméthylbenzimidazole, mais une autre base, par exemple, le 5-hydroxybenzimidazole, pour la vitamine B<sub>12</sub>-facteur III synthétisée entre autre par Clostridium thermoaceticum et Methanosarcina barkeri (Iron et al., 1984). Ces similitudes structurales expliquent que les voies métaboliques de biosynthèse des 30 cobalamines et des cobamides soient communes pour la majeure partie.

Les cobalamines sont synthétisées presque exclusivement par des bactéries, selon un processus complexe et encore très mal connu, qui peut se diviser en quatre étapes (figure 2) :

i) synthèse de l'uroporphyrinogène III (ou uro'gen III),  
35 puis

ii) transformation de l'uro'gen III en acide cobyrinique suivie de

iii) transformation de celui-ci en cobinamide et

iv) construction de la boucle nucléotidique inférieure  
5 avec incorporation de la base particulière (5,6-diméthylbenzimidazole dans le cas des cobalamines).

Pour le coenzyme B<sub>12</sub>, il est probable que l'addition du groupement 5'-déoxy-5'-adénosyl se produit peu après que le noyau corrine soit synthétisé (Huennekens et al., 1892).

10 Dans le cas des cobamides, seule l'étape de synthèse et d'incorporation de la base inférieure est différente.

La première partie de la biosynthèse des cobalamines est très bien connue puisqu'elle est commune à celle des hèmes ainsi qu'à celle des chlorophylles (Battersby et al., 1980). Elle fait intervenir  
15 successivement la δ-aminolevulinate synthase (EC 2.3.137), la δ-aminolévulinate déhydrase (EC 4.2.1.24), la porphobilinogène déaminase (EC 4.3.1.8) et l'uro'gen III cosynthase (EC 4.2.1.75) qui transforment le succinyl CoA et la glycine en uro'gen III. Toutefois la première étape se fait chez certains organismes [par exemple E.coli 20 (Avissar et al., 1989) et chez les bactéries méthanogènes (Kannangara et al., 1989), par exemple] par la conversion grâce à un complexe multienzymatique de l'acide glutamique en acide δ-aminolévulinique.

Entre l'uro'gen III et l'acide cobyrinique, seuls trois dérivés d'intermédiaires ont été purifiés à ce jour ; il s'agit des  
25 Facteurs FI, FII et FIII qui sont des produits d'oxydation respectivement des trois intermédiaires précorrine-1, précorrine-2 et précorrine-3 qui correspondent aux dérivés mono-, di- et triméthylés de l'uro'gen III (figure 3) ; ces intermédiaires sont obtenus par des transferts successifs de groupements méthyl à partir du SAM 30 (S-adénosyl-L-méthionine) sur l'uro'gen III aux positions C2, C7 et C20 respectivement. Les autres réactions qui ont lieu pour donner l'acide cobyrinique sont, outre cinq autres transferts de groupements méthyl à partir du SAM en C17, C12, C1, C15 et C5, l'élimination du carbone en C20, la décarboxylation en C12 et l'insertion d'un atome de

cobalt (figure 4). Ces étapes de biosynthèse ont été déduites à partir d'expériences effectuées *in vitro* sur des extraits acellulaires de Propionibacterium shermanii ou de Clostridium tetanomorphum. Dans ces extraits, l'acide cobyrinique est obtenu par transformation de 5 l'uro'gen III, après incubation dans des conditions appropriées en anaérobiose (Batterby et al., 1982). Aucun intermédiaire entre le précorrine-3 et l'acide cobyrinique, pouvant être transformé en corrinoïdes par incubations ultérieures avec des extraits de bactéries productrices de cobalamines, n'a été isolé à ce jour chez ces 10 microorganismes. La difficulté d'isoler et d'identifier ces intermédiaires est liée à

- i) leur grande instabilité
- ii) leur sensibilité à l'oxygène et
- iii) leur faible niveau d'accumulation *in vivo*.

15 Dans cette partie de la voie, seule une enzyme de Pseudomonas denitrificans a été purifiée et étudiée ; il s'agit de la SAM:uro'gen III méthyltransférase (Blanche et al., 1989) appelée SUMT.

Entre l'acide cobyrinique et le cobinamide, les réactions suivantes sont effectuées :

20 i) addition du groupement 5'-désoxyadénosyl (s'il s'agit du coenzyme B<sub>12</sub> qui doit être synthétisé),  
ii) amidation de six des sept fonctions carboxyliques par addition de groupements aminés,  
iii) amidation de la dernière fonction carboxylique  
25 (chaîne d'acide propionique du noyau pyrrole D) par addition du R-1-amino-2-propanol (figure 2).

Il n'a pas été élucidé s'il existait réellement un ordre dans les amidations (Herbert et al., 1970). Enfin, aucun dosage d'activité dans cette partie de la voie n'a été décrit sauf en ce qui 30 concerne l'addition du groupement 5'-désoxyadénosyl (Huennekens et al., 1982).

La dernière étape de la biosynthèse d'une cobalamine, par exemple le coenzyme B<sub>12</sub>, comprend quatre phases successives décrites sur la figure 5 (Huennekens et al., 1982), à savoir :

35 i) la phosphorylation du groupement hydroxyl du résidu

aminopropanol du cobinamide pour donner le cobinamide phosphate, puis  
ii) l'addition d'une guanosine diphosphate par réaction  
avec la guanosine 5'-triphosphate ; le composé obtenu est le  
GDP-cobinamide (Friedmann, 1975) qui  
5       iii) réagit avec le α-ribazole 5'-phosphate, lui-même  
synthétisé à partir de la riboflavine, pour donner l'adénosyl-  
cobalamine 5'-phosphate (Friedmann et al., 1968) qui  
iv). par déphosphorylation conduit au coenzyme B<sub>12</sub>  
(Schneider et Friedmann, 1972).

10           Parmi les bactéries capables de produire des cobalamines,  
on peut citer notamment :

15           Agrobacterium tumefaciens,  
Agrobacterium radiobacter  
Bacillus megaterium  
15           Clostridium sticklandii  
Clostridium tetanomorphum  
Clostridium thermoaceticum  
20           Corynebacterium XG  
Eubacterium limosum  
20           Methanobacterium arbophilicum  
Methanobacterium ivanovii  
Methanobacterium ruminantium  
Methanobacterium thermoautotrophicum  
25           Methanosarcina barkeri  
Propionobacterium shermanii  
Protaminobacter ruber  
Pseudomonas denitrificans  
Pseudomonas putida  
30           Rhodopseudomonas sphaeroides  
Salmonella typhimurium  
Spirulina platensis  
Streptomyces antibioticus  
Streptomyces aureofaciens

Streptomyces griseusStreptomyces olivaceus

Au niveau industriel, en raison de la grande complexité des mécanismes de biosynthèse, la production des cobalamines, et en particulier de la vitamine  $B_{12}$ , est exclusivement microbiologique. Elle est réalisée par des cultures en grands volumes des bactéries Pseudomonas denitrificans, Propionobacterium shermanii et Propionibacterium freudenreichii (Florent, 1986). Les souches utilisées pour la production industrielle sont issues de souches sauvages; elles peuvent avoir subi de nombreux cycles de mutation au hasard puis de sélection de clones améliorés pour la production de cobalamines (Florent, 1986). Les mutations sont obtenues par mutagénèse avec des agents mutagènes ou par traitements physiques tels que des traitements aux rayons ultra-violets (Barrère et al., 1981). Par cette méthode empirique, des mutations au hasard sont obtenues et améliorent la production de cobalamines. Par exemple, il est décrit qu'à partir de la souche originale de Pseudomonas denitrificans initialement isolée par Miller et Rosenblum (1960, brevet US 2 938 822) la production de ce microorganisme a été graduellement augmentée en dix ans, par les techniques citées ci-dessus de 0,6 mg/l à 60 mg/l (Florent, 1986). Pour les bactéries du genre Propionibacterium, [Propionobacterium shermanii (ATCC 13673) et freudenreichii (ATCC 6207)] les mêmes valeurs de production semblent être décrites dans la littérature ; par exemple une production de 65 mg/l a été décrite (brevet Européen 87920). Toutefois, aucun crible n'a encore été décrit permettant de sélectionner ou de repérer facilement soit des mutants surproducteurs de cobalamines soit des mutants nettement améliorés dans leur production de cobalamines.

Au niveau génétique, peu de travaux ont été effectués à ce jour. Le clonage de gènes cob (codant pour des enzymes impliquées dans le procédé de biosynthèse) a été décrit chez Bacillus megaterium (Brey et al., 1986). Onze groupes de complémentation ont été identifiés par complémentation de mutants cob de Bacillus megaterium avec des plasmides portant différents fragments d'ADN de Bacillus megaterium.

Ces gènes sont groupés sur le même locus, porté par un fragment de 12 kb.

Des études ont également été menées sur les gènes cob de Salmonella typhimurium. Sans que le clonage de ceux-ci ait été décrit, 5 il a été montré que presque tous les gènes de biosynthèse des cobalamines sont regroupés entre les minutes 40 et 42 du chromosome (Jeter et Roth, 1987). Seul le locus cysG, qui doit permettre la transformation de l'uro'gen III en précorrine-2 ne fait pas partie de ce groupe de gènes. Toutefois, l'activité codée par ce locus, ainsi 10 que ses propriétés biochimiques n'ont pas été décrites.

En outre, des phénotypes ont été associés aux mutations cob. Chez Salmonella typhimurium et chez Bacillus megaterium, les 15 mutants cob ne montrent plus de croissance sur milieu minimum avec de l'éthanolamine comme source de carbone ou comme source d'azote (Roof et Roth, 1988). Ceci est dû au fait qu'une enzyme du catabolisme de l'éthanolamine, l'éthanolamine ammonia-lyase (EC 4.3.1.7), a pour cofacteur le coenzyme B<sub>12</sub>; les mutants cob ne synthétisant plus de coenzyme B<sub>12</sub>, ils ne peuvent plus croître avec l'éthanolamine comme 20 source de carbone et/ou comme source d'azote. Des mutants metE de Salmonella typhimurium n'ont plus qu'une homocystéine-méthyl-transférase (EC 2.1.1.13) méthylcobalamine dépendante. Les mutants cob de Salmonella typhimurium metE sont auxotropes pour la méthionine (Jeter et al., 1984).

Chez Pseudomonas denitrificans et Agrobacterium tumefaciens 25 des phénotypes associés à une déficience totale en synthèse de cobalamines n'ont pas été décrits à ce jour.

Enfin, des travaux sur Pseudomonas denitrificans (Cameron et al., 1989) ont abouti au clonage de fragments d'ADN portant des gènes cob de cette bactérie. Ceux-ci sont répartis en quatre groupes 30 de complémentation portés par au moins 30 kb d'ADN. Au moins quatorze groupes de complémentation ont été identifiés par complémentation hétérologue de mutants cob d'Agrobacterium tumefaciens et de Pseudomonas putida avec des fragments d'ADN de Pseudomonas denitrificans portant des gènes cob.

35 Cependant, jusqu'à maintenant, aucun de ces gènes n'a été

purifié, et aucune séquence nucléotidique n'a été décrite. De même, aucune identification de protéine, ni aucune fonction catalytique attribuée au produit de ces gènes n'ont été décrites. De plus, aucune amélioration de production de cobalamines par les techniques d'ADN recombinant n'a pu être obtenue. L'amplification de gènes cob de Bacillus megaterium ne procure pas, chez la souche à partir de laquelle ils ont été clonés, une amélioration de production de cobalamines (Brey et al., 1986). Chez Salmonella typhimurium, des études physiologiques ont été menées afin de déterminer des conditions dans lesquelles une forte transcription des gènes cob étudiés était observée (Escalante et Roth, 1987). Dans ces conditions, il n'y a pas d'amélioration de la production de cobalamines, bien que des gènes de la voie de biosynthèse soient plus exprimés que dans les conditions standard de culture.

La présente invention résulte de l'identification précise de séquences d'ADN codant pour des polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides. Un objet de l'invention concerne donc les séquences d'ADN codant pour les polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides. Plus particulièrement, l'invention a pour objet les gènes cobA, cobB, cobC, cobD, cobE, cobF, cobG, cobH, cobI, cobJ, cobK, cobL, cobM, cobN, cobO, cobP, cobQ, cobS, cobT, cobU, cobV, cobW, cobX et corA, toute séquence d'ADN homologue de ces gènes résultant de la dégénérescence du code génétique, ainsi que les séquences d'ADN, de toute origine (naturelle, synthétique, recombinante) qui hybrident et/ou qui présentent des homologies significatives avec ces séquences ou avec des fragments de celles-ci et qui codent pour des polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides. L'invention a aussi pour objet les gènes contenant ces séquences d'ADN.

Les séquences d'ADN selon la présente invention ont été isolées à partir de différentes souches : une souche industrielle, Pseudomonas denitrificans SC510, dérivée de la souche MB580 (brevet US 3 018 225), par complémentation de mutants cob de A.tumefaciens et P.putida; et de Methanobacterium ivanovii. Les clones obtenus ont pu

être analysés précisément, notamment par cartographie à l'aide d'insertions d'un dérivé du transposon Tn5. Ces études génétiques ont permis de localiser les gènes cob ou cor sur la carte de restriction et de réaliser leur séquençage. Une analyse des phases ouvertes a 5 ensuite permis de mettre en évidence les régions codantes de ces fragments d'ADN.

La présente invention a aussi pour objet l'utilisation de ces séquences nucléotidiques pour cloner les gènes cob d'autres bactéries. En effet il est connu que pour des protéines catalysant les 10 mêmes activités, les séquences sont conservées avec, comme divergence, la divergence évolutive (Wein-Hsiung et al., 1985). Il est montré dans la présente invention qu'il y a une homologie significative entre les séquences nucléotidiques de gènes de différents microorganismes codant pour des polypeptides impliqués dans la biosynthèse des cobalamines 15 et/ou des cobamides. Les différences qui apparaissent résultent de la dégénérescence évolutive, et de la dégénérescence du code génétique qui est liée au pourcentage en GC du génome du microorganisme étudié (Wein-Hsiung et al., 1985).

Selon la présente invention, une sonde peut être faite 20 avec une ou plusieurs des séquences d'ADN de Pseudomonas denitrificans notamment, ou avec des fragments de celles-ci, ou avec des séquences analogues présentant un degré de dégénérescence spécifique, au niveau de l'usage des codons et du pourcentage en GC de l'ADN de la bactérie que l'on veut étudier. Dans ces conditions, il est possible de 25 détecter un signal d'hybridation spécifique entre la sonde et des fragments d'ADN génomique de la bactérie étudiée ; ce signal d'hybridation spécifique correspond à l'hybridation de la sonde avec les gènes cob isofonctionnels de la bactérie. Les gènes cob ainsi que leurs produits peuvent ensuite être isolés, purifiés et caractérisés. 30 L'invention fournit ainsi un moyen permettant, par hybridation, d'accéder aux séquences nucléotidiques et aux polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides de tout microorganisme.

La présente invention a également pour objet un ADN 35 recombinant contenant au moins une séquence d'ADN codant pour un

polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides, et notamment, un ADN recombinant dans lequel la ou lesdites séquences sont placées sous le contrôle de signaux d'expression.

A cet égard, on peut en particulier positionner en 5' de la séquence d'ADN des régions promotrices. De telles régions peuvent être homologues ou hétérologues de la séquence d'ADN. En particulier, des promoteurs bactériens forts, tels que le promoteur de l'opéron tryptophane P<sub>trp</sub> ou de l'opéron lactose P<sub>lac</sub> de E.coli, le promoteur gauche ou droit du bactériophage lambda, les promoteurs forts de phages de bactéries, telles que les corynebactéries, les promoteurs fonctionnels chez les bactéries gram-négatives, tel que le promoteur P<sub>tac</sub> de E.coli, le promoteur P<sub>xylS</sub> des gènes du catabolisme du xylène du plasmide TOL, le promoteur de l'amylase de Bacillus subtilis P<sub>amy</sub>, pourront être utilisés. On peut citer également les promoteurs dérivés de gènes glycolytiques de levures, tels que les promoteurs des gènes codant pour le phosphoglycérate kinase, la glycéraldéhyde-3-phosphate déshydrogénase, la lactase ou l'éholase, qui pourront être utilisés lorsque l'ADN recombinant sera introduit dans un hôte eucaryote. Un site de fixation des ribosomes sera également positionné en 5' de la séquence d'ADN et il pourra être homologue ou hétérologue, tel le site de fixation des ribosomes du gène CII du bactériophage lambda.

Des signaux nécessaires à la terminaison de la transcription pourront être placés en 3' de la séquence d'ADN.

L'ADN recombinant selon la présente invention peut ensuite être introduit directement dans une cellule hôte compatible avec les signaux d'expression choisis ou être cloné sur un vecteur plasmidique pour permettre d'introduire de manière stable la séquence d'ADN en question dans la cellule hôte.

Un autre objet de l'invention concerne les plasmides ainsi obtenus contenant une séquence d'ADN codant pour un polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides. Plus précisément, ces plasmides contiennent aussi un système de réPLICATION fonctionnel et un marqueur de sélection.

L'invention a aussi pour objet les cellules hôtes dans lesquelles une ou plusieurs séquences d'ADN telles que définies

précédemment, ou un plasmide tel que défini ci-dessus, a été introduit.

Un autre objet de l'invention concerne un procédé de production de polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides. Selon ce procédé, on introduit dans une cellule hôte une séquence d'ADN telle que décrite précédemment, on cultive cette cellule recombinante dans des conditions d'expression de ladite séquence, puis on récupère les polypeptides produits.

Les cellules hôtes qui pourront être utilisées dans ce but sont aussi bien des procaryotes que des eucaryotes, des cellules animales ou des cellules végétales. Préférentiellement, elles seront choisies parmi les bactéries et, en particulier, les bactéries du genre E.coli, P.denitrificans, A.tumefaciens ou R.meliloti.

Une autre utilisation des séquences d'ADN selon la présente invention réside dans un procédé d'amplification de la production de cobalamines et/ou de cobamides ou de leurs précurseurs de biosynthèse, par les techniques d'ADN recombinant. En effet, si la limitation du flux métabolique de la biosynthèse des cobalamines et/ou des cobamides ou de leurs précurseurs est due à une limitation dans l'activité d'une enzyme dans la voie de biosynthèse, l'augmentation de cette activité par augmentation de l'expression de cette même enzyme à l'aide des techniques d'ADN recombinant (amplification génique, substitution des signaux de transcription-traduction par des signaux plus efficaces ...) conduira à une augmentation de la biosynthèse des cobalamines et/ou des cobamides. Il est possible également que la limitation de la production de cobalamines et/ou de cobamides résulte d'une régulation biochimique. Dans ce cas, le ou les gènes cob correspondant à l'enzyme régulée pourront être mutagénisés in vitro spécifiquement afin d'obtenir des gènes mutés dont les produits auront perdu les régulations allant à l'encontre d'une amélioration de la production.

Le procédé selon la présente invention consiste à introduire dans un microorganisme producteur de cobalamines et/ou de cobamides, ou seulement potentiellement producteur de ces composés (c'est-à-dire déficient dans une ou plusieurs des étapes de la biosynthèse), une séquence d'ADN telle que définie plus haut, puis à

cultiver ce microorganisme dans des conditions d'expression de ladite séquence et de synthèse des cobalamines et/ou des cobamides, et enfin à récupérer les cobalamines et/ou les cobamides produits. Un tel procédé est applicable en particulier à tous les microorganismes producteurs cités page 4, et plus spécifiquement aux microorganismes du genre P. denitrificans, Rhizobium melitoti, ou Agrobactérium tumefaciens. Dans un mode de réalisation préféré, le microorganisme est P. denitrificans, et en particulier, la souche SC510. Concernant les microorganismes potentiellement producteurs, les séquences d'ADN utilisées seront celles correspondant aux étapes de la biosynthèse que le microorganisme ne peut réaliser.

A l'aide de la présente invention et par les diverses stratégies exposées ci-dessus, une amélioration de la production de cobalamines et/ou de cobamides ou de leurs précurseurs pour tout microorganisme producteur, ou potentiellement producteur, de cobalamines et/ou de cobamides, pourra être obtenue. Il suffira de cultiver ce microorganisme recombiné dans des conditions appropriées pour la production de cobalamines et pour l'expression des séquences d'ADN introduites. Cette culture pourra se faire en batch ou bien en continu, et la purification des cobalamines pourra se faire par les méthodes déjà utilisées au niveau industriel (Florent, 1986). Ces méthodes comprennent entre autre :

i) la solubilisation des cobalamines et leur conversion en leur forme cyano (par exemple en traitant le moût de fermentation à la chaleur, avec du cyanure de potassium en présence de nitrite de sodium), puis

ii) la purification des cyanocobalamines en différentes étapes qui peuvent être par exemple

a) l'adsorption sur différents substrats tels que l'amberlite IRC50, le Dowex 1X2 ou l'amberlite XAD 2 suivie d'une élution avec un mélange eau-alcool ou eau phénol, puis

b) l'extraction dans un solvant organique et enfin

c) la précipitation ou cristallisation à partir de la phase organique soit par addition de réactifs ou dilution dans un solvant approprié, soit par évaporation.

La présente invention montre de plus qu'il est possible, par les techniques d'ADN recombinant, d'améliorer la production de cobalamines d'une bactérie productrice de cobalamines en cumulant des améliorations. Ceci revient à obtenir une première amélioration comme cela est décrit ci-dessus puis à améliorer cette amélioration toujours à l'aide des techniques d'ADN recombinant, c'est-à-dire par exemple en amplifiant des gènes de biosynthèse des cobalamines.

Un autre objet de la présente invention concerne les polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides. En particulier, la présente invention a pour objet tous les polypeptides, ou des dérivés ou des fragments de ces polypeptides, qui sont codés par les séquences d'ADN décrites précédemment, et qui sont impliqués dans la voie de biosynthèse des cobalamines et/ou des cobamides. La séquence en acides aminés de ces polypeptides est décrite, ainsi que certaines de leurs caractéristiques physicochimiques. Une activité enzymatique ou des propriétés spécifiques ont également été associées à chacun d'entre-eux.

A cet égard, l'invention a pour objet les polypeptides intervenant dans la transformation du précorrine-3 en acide cobyrinique a,c-diamide, et plus particulièrement dans le transfert d'un groupement méthyl du SAM aux positions C1, C5, C11, C15, et C17.

L'invention a aussi pour objet les polypeptide:

. intervenant dans la transformation de l'acide cobyrique en cobinamide, ou

25 . possédant une activité S-adénosyl-L-méthionine:précorrine-2 méthyl transférase (SP2MT), ou

. possédant une activité acide cobyrinique et/ou hydrogénobyrinique a,c-diamide synthase, ou

. possédant une activité précorrin-8x mutase, ou

30 . possédant une activité nicotinate-nucléotide: dimethylbenzimidazole phosphoribosyltransferase, ou

. possédant une activité cobalamine-(5'-phosphate) synthase, ou

. possédant une activité cobyric acid synthase, ou

35 . possédant une activité cob(I)alamin adénosyltransférase,

ou

- . possédant une activité precorrin-6x réductase, ou
- . intervenant dans la transformation de l'acide hydrogénobyrinique a,c-diamide en acide cobyrinique a,c-diamide.

5 Avantageusement, l'invention a pour objet un polypeptide choisi parmi les protéines COBA, COBB, COBC, COBD, COBE, COBF, COBG, COBH, COBI, COBJ, COBK, COBL, COBM, COBN, COBO, COBP, COBQ, COBS, COBT, COBU, COBV, COBW, COBX et CORA présentées aux figures 15, 16, 40, 41 et 47.

10 De plus, l'utilisation des sondes d'hybridation décrites précédemment permet, à partir de gènes isolés dans d'autres microorganismes, de caractériser et d'isoler les polypeptides isofonctionnels d'autres microorganismes. De cette manière, la présente invention montre que la séquence d'une protéine COB de  
15 Pseudomonas denitrificans est significativement homologue aux séquences de protéines d'autres microorganismes présentant le même type d'activité. Entre ces protéines COB catalysant la même réaction chez des microorganismes différents, seules les distances évolutives ont introduit des variations (Wein-Hsiung et al., 1985). La présente  
20 invention à également pour objet ces polypeptides isofonctionnels.

L'attribution d'une activité enzymatique particulière est le résultat d'une analyse qui peut être effectuée selon diverses stratégies. En particulier, des études d'affinité in vitro, vis-à-vis du SAM (S-adénosyl-L-méthionine), permettent d'attribuer à une  
25 protéine capable de fixer le SAM, une activité méthyltransférase, et donc son implication dans une des étapes de transfert des groupements méthyl qui se produisent entre l'uro'gen III et l'acide cobyrinique. Un autre moyen d'apprécier l'activité de ces polypeptides consiste à doser les intermédiaires de la voie de biosynthèse des cobalamines qui  
30 sont accumulés chez des mutants incapables d'exprimer ces polypeptides (identifiés par des expériences de complémentation). Ces analyses permettent de déduire que le polypeptide en question a pour substrat l'intermédiaire accumulé, ce qui permet de situer et de définir son

activité dans la voie de biosynthèse. La présente invention décrit également un procédé de dosage des activités enzymatiques de la voie de biosynthèse, applicable à toute souche productrice de cobalamines et/ou de cobamides. Ces dosages permettent de purifier à partir de 5 toute souche productrice de ces composés, l'activité enzymatique dosée. A partir de cette activité purifiée, la séquence NH<sub>2</sub>-terminale de la protéine COB en question, ou bien celle des sous-unités de cette protéine, peut être effectuée, ce qui permet d'identifier le ou les gènes de structure qui codent pour l'activité en question. Pour 10 Pseudomonas denitrificans, les gènes de structure qui codent pour des activités de la voie de biosynthèse sont identifiés en trouvant pour chaque séquence NH<sub>2</sub>-terminale, la protéine COB ayant la même séquence NH<sub>2</sub>-terminale.

La présente invention décrit aussi un procédé permettant 15 l'identification et le dosage, chez des souches productrices de cobalamines ou des mutants non producteurs, d'intermédiaires de la voie de biosynthèse des cobalamines ou d'autres corrinoïdes. Ces intermédiaires peuvent être dosés aussi bien dans des moûts de culture que dans les cellules elles-mêmes. Les intermédiaires qui peuvent être 20 dosés sont tous les corrinoïdes qui se trouvent dans la voie de biosynthèse après l'acide cobyrinique à savoir outre l'acide cobyrinique, l'acide cobyrinique monoamide, l'acide cobyrinique diamide, l'acide cobyrinique triamide, l'acide cobyrinique tétraamide, l'acide cobyrinique pentaamide, l'acide cobyrinique, le cobinamide, le 25 cobinamide phosphate, le GDP-cobinamide, le coenzyme B<sub>12</sub>-phosphate et le coenzyme B<sub>12</sub>. Les formes cyano et coenzyme de ces produits peuvent être dosées par cette technique.

D'autres objets et avantages de la présente invention apparaîtront à la lecture des exemples et des dessins suivants, qui 30 doivent être considérés comme illustratifs et non limitatifs.

#### DEFINITION DES TERMES EMPLOYES ET ABREVIATIONS

ACDAS: acide cobyrinique a,c-diamide synthase

ADN recombinant: ensemble de techniques qui permettent soit d'associer

au sein du même microorganisme des séquences d'ADN qui ne le sont pas naturellement, soit de mutagéniser spécifiquement un fragment d'ADN.

ATP: adénosine 5'-triphosphate

BSA: sérum albumine bovine

5 CLHP: Chromatographie liquide à haute performance.

cluster: groupe de gènes.

Cob: correspond au phénotype de niveau réduit (au moins 10 fois moins que le contrôle) de production de cobalamines.

Codon stop: codon de terminaison de traduction

10 Corrinoïdes : dérivés de l'acide cobyriénique possédant le noyau corrine.

dGTP: 2'-désoxyguanosine 5'-triphosphate

DMBI: diméthylbenzimidazole

dNTP: 2'-désoxyribonucléosides 5'-triphosphates

15 DTT: dithiothréitol

gène cob: gène impliqué dans la biosynthèse du cobinamide à partir de l'uro'gen III.

gène cor: gène impliqué dans la biosynthèse des corrinoïdes à partir de l'uro'gen III.

20 kb: kilobases

NN:DMBI PRT :

ORF: phase ouverte

pb: paires de bases

Protéine COB: protéine intervenant soit comme catalyseur dans la voie de biosynthèse des cobalamines, soit comme protéine régulatrice dans le réseau de régulation des gènes cob, soit les deux.

Protéine COR: protéine intervenant soit comme catalyseur dans la voie de biosynthèse des corrinoïdes, soit comme protéine régulatrice dans le réseau de régulation des gènes cor, soit les deux.

30 SAM: S-adénosyl-L-méthionine

SDS: sodium dodécyl sulfate

SP2MT: SAM-L-méthionine:précorrine-2 méthyltransférase

SUMT: SAM:uro'gen III méthyltransférase

Uro'gen III: uroporphyrinogène III.

**Légendes des figures:****Figure 1:**

Structure du coenzyme B12; le groupement 5'-déoxyadénosyl est remplacé par un groupement CH<sub>3</sub> pour la méthylcobalamine, par un groupement 5 cyano pour la cyanocobalamine, par un groupement hydroxyl pour l'hydroxocobalamine.

**Figure 2:**

Biosynthèse des cobalamines et différentes étapes de cette biosynthèse suivant la littérature. X: ligands axiaux du cobalt; le ligand en a peut être différent du ligand en b. R: ligand en a du cobalt qui définit le type de cobalamine (voir figure 1).

**Figure 3:**

Structures de l'uro'gen III, du précorrine-1, du précorrine-2 et du précorrine-3.

**Figure 4:**

Formules développées de l'uro'gen III et de l'acide cobyrinique. D'après la littérature, entre l'uro'gen III et l'acide cobyrinique il se produit 8 transferts de méthyles SAM-dépendants successivement en C<sub>2</sub>, C<sub>7</sub>, C<sub>20</sub>, C<sub>17</sub>, C<sub>12</sub>, C<sub>1</sub>, C<sub>15</sub> et C<sub>5</sub>, une décarboxylation en C<sub>12</sub>, l'élimination du carbone en C<sub>20</sub> et l'insertion de l'atome de cobalt. X: ligands axiaux du cobalt; le ligand en a peut être différent du ligand en b.

**Figure 5:**

Dernières étapes de la biosynthèse des cobalamines. Afin de clarifier le schéma des détails du noyau corrine ont été omis. Les cinq étapes enzymatiques sont représentées: 1, cobinamide kinase; 2, cobinamide phosphate guanylyltransférase; 3, cobalamine 5'-phosphate synthase; 4, cobalamine 5'-phosphate phosphohydrolase; 5, nicotinate nucléotide:DMBII phosphoribosyltransférase.

**Figure 6:** Cartes de restriction des fragments ClaI-HindIII-HindIII-HindIII de 5.4 kb; EcoRI de 8.7 kb; SalI-SalI-SalI-SalI-SalI-BglI de 4748 pb et SstI-SstI-BamHI de 3855 pb . Ne figurent que les 20 enzymes de restriction qui coupent le moins fréquemment l'ADN. Les sites de coupure de chaque

enzyme sont indiqués par un trait vertical.

Figure 7:

Séquence nucléotidique des deux brins du fragment ClaI-HindIII-HindIII-HindIII de 5378 pb de *Pseudomonas denitrificans*.

5 Le brin situé en haut est à lire 5' vers 3' dans le sens gauche droite qui correspond à l'orientation gauche-droite du fragment de la carte de restriction présentée à la figure 6. Le site ClaI se trouve à la position 23 (début du site de coupure) car sur cette séquence se trouvent des sites de restriction PstI, SalI et XbaI qui sont apparus  
10 lors des clonages dans des multisites en vue du séquençage. La séquence du fragment ClaI-HindIII-HindIII-HindIII commence donc à la position 23.

Figure 8:

Séquence nucléotidique des deux brins du fragment EcoRI de 8753 pb de *Pseudomonas denitrificans*. Le brin situé en haut est à lire 5' vers 3' dans le sens gauche-droite qui correspond à l'orientation gauche-droite du fragment de la carte de restriction présentée à la figure 6.

Figure 9:

20 Analyse des probabilités des phases codantes d'après usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur les 6 phases de lecture de la séquence du fragment ClaI-HindIII-HindIII-HindIII de 5378 pb. Pour les phases qui appartiennent au même brin codant, la phase la plus probable  
25 correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase.

1. Séquence allant du nucléotide 1 au nucléotide 1200. Grâce à cette analyse, la phase ouverte 1 est identifiée. Elle commence à l'ATG en position 549 et se termine au TGA en position  
30 1011.

2. Séquence allant du nucléotide 1000 au nucléotide 2200. Grâce à cette analyse, la phase ouverte 2 est identifiée. Elle commence à l'ATG en position 1141 et se termine au TGA en position 1981.

35 3. Séquence allant du nucléotide 1800 au nucléotide 3400.

Grâce à cette analyse, la phase ouverte 3 est identifiée. Elle commence à l'ATG en position 1980 et se termine au TGA en position 3282.

4. Séquence allant du nucléotide 3000 au nucléotide 4500.  
5 Grâce à cette analyse, la phase ouverte ouverte 4 est identifiée. Elle commence à l'ATG en position 3281 et se termine au TGA en position 4280.

5. Séquence allant du nucléotide 3800 au nucléotide 5378.  
Grâce à cette analyse, la phase ouverte 5 est identifiée. Elle  
commence au GTG en position 4284 et se termine au TGA en position  
5253.

Figure 10:  
Analyse des probabilités des phases codantes d'après usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur les 6 phases de lecture du fragment EcoRI de 8753 pb. Pour les phases qui appartiennent au même brin codant, la phase la plus probable correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase.

1. Séquence allant du nucléotide 650 au nucléotide 1650.  
20 Grâce à cette analyse, la phase ouverte 6 est identifiée. Elle commence à l'ATG en position 736 et se termine au TAG en position 1519.

2. Séquence allant du nucléotide 1400 au nucléotide 3100.  
Grâce à cette analyse, la phase ouverte 7 est identifiée. Elle  
commence à l'ATG en position 1620 et se termine au TAG en position 2997.

3. Séquence allant du nucléotide 2700 au nucléotide 3700.  
Grâce à cette analyse, la phase ouverte 8 est identifiée. Elle  
commence à l'ATG en position 3002 et se termine au TGA en position  
30 3632.

4. Séquence allant du nucléotide 3500 au nucléotide 4100.  
Grâce à cette analyse, la phase ouverte 9 est identifiée. Elle  
commence au GTG en position 3631 et se termine au TGA en position  
4366.

35 5. Séquence allant du nucléotide 4150 au nucléotide 5150.

Grâce à cette analyse, la phase ouverte 10 est identifiée. Elle commence à l'ATG en position 4365 et se termine au TGA en position 5127.

6. Séquence allant du nucléotide 5000 au nucléotide 6000.  
5 Grâce à cette analyse, la phase ouverte 11 est identifiée. Elle commence à l'ATG en position 5893 et se termine au TAG en position 5110.

7. Séquence allant du nucléotide 5700 au nucléotide 7200.  
Grâce à cette analyse, la phase 12 est identifiée. Elle commence à  
10 l'ATG en position 5862 et se termine au TAA en position 7101.

8. Séquence allant du nucléotide 7000 au nucléotide 8000.  
Grâce à cette analyse, la phase ouverte 13 est identifiée. Elle commence à l'ATG en position 7172 et se termine au TTG en position 7931.

15 Figure 11:  
Construction des plasmides pXL556, pXL545 et pXL723. Un fragment ClaI-EcoRV de 2.4 kb contenant les gènes cobA et cobE est excisé du fragment de 5.4 kb puis purifié. Un "linker" EcoRI est ajouté au site EcoRV, puis le fragment est inséré dans le pXL59 entre les sites  
20 ClaI-EcoRI. Le plasmide ainsi construit est nommé pXL556. La construction est comparable pour le pXL545 : un fragment ClaI-HindIII-HindIII de 1.9 kb est excisé du fragment de 5.4 kb puis purifié. Ce fragment contient uniquement le gène CobE. Un "linker" EcoRI est ajouté au site HindIII, puis le fragment est inséré dans le  
25 pXL59 entre les sites ClaI-EcoRI.

Le pXL723 est construit comme suit: un fragment EcoRI-HindIII de 2.3 kb est excisé du fragment de 5.4 kb, purifié, puis les extrémités sont remplies par le grand fragment de l'ADN polymérase I de E. coli. Ce fragment est cloné dans le pRK290 (Ditta et al., 1981) digéré par  
30 EcoRI puis traité avec le large fragment de l'ADN polymérase I de E. coli afin de remplir les extrémités. Les sites de restrictions qui sont figurés entre parenthèse correspondent à des sites qui ont disparu après traitement avec le large fragment de l'ADN polymérase I de E. coli.

35 1 , fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2,

fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); 3, fragment BamHI-SstI contenant l'opéron lactose de E. coli sans son promoteur, l'opérateur, le site d'initiation de la traduction et les 8 premiers codons non essentiels de lacZ (Casadaban et al., 1983); 4, 5 fragment Sau3AI de Pseudomonas putida KT2440 (Bagdasarian et al., 1981); ori, origine de réPLICATION; nic, site de relaxation; mob, locus essentiel pour la mobilisation; Kmr gén-e de résistance à la kanamycine (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI.

10 Figure 12:

Etudes des insertions des transposons Tn5Spr et Tn5 sur le fragment de 5378 bp. Les insertions du transposon Tn5 sur le plasmide pXL723 sont figurées comme à la figure 14; celles du transposon Tn5Spr, dans le chromosome de la souche SBL27 Rifr, sont encadrées; les insertions 15 dans le chromosome de SC510 Rifr de cassettes portant le gène de résistance à la kanamycine (1630 et 1631) sont représentées avec une flèche, selon l'orientation de la transcription du gène de résistance à la kanamycine, sous le numéro de l'insertion. Les phases ouvertes déduites de la séquence sont portées sur cette figure (de cobA à 20 cobE); des signes + ou - sont représentés sous chaque insertion de transposon ou de cassette de résistance pour indiquer que l'insertion est inactivationnelle (-) ou non (+) soit pour la complémentation de différents mutants (cas des insertions de transposons Tn5) ou que l'insertion abolit la production de cobalamines de la souche dans 25 laquelle elle a lieu. Il y a une absence de complémentation lorsque le mutant recombiné synthétise moins de trois fois moins de cobalamines que le niveau de synthèse de la souche dont le mutant dérive. Les inserts des plasmides pXL545Ω, pXL1500, pXL1397 et pXL302 sont représentés avec les sites de restriction se trouvant à leurs 30 extrémités. Ces inserts sont clonés dans les plasmides à large spectre d'hôte pXL435 et pXL59 (Cameron et al., 1989).

Le plasmide pXL545Ω correspond au plasmide pXL545 décrit à la figure 11 avec en plus le fragment BamHI de 2 kb du pHp45Ω (Prentki et Krisch) contenant un gène de résistance à la spectinomycine cloné au 35 site BamHI du pXL545.

Le plasmide pXL1500 correspond au fragment BglIII-SstI de 4.2 kb présenté sur cette figure cloné aux sites BamHI et SstI du pKT230 (Bagdasarian et al., 1981) présenté à la figure 30; le pXL1397 correspond au fragment HindIII-SstI de 2.4 kb désigné sur la figure 5 inséré entre les sites HindIII et SstI du multisite du pXL435 (Cameron et al., 1989) décrit sur la figure 30; le plasmide pXL302 correspond au fragment EcoRI-HindIII de 2.3 kb tel qu'il est décrit sur la figure insérée entre les sites EcoRI et HindIII du pXL59 (Cameron et al., 1989), décrit sur la figure 30, le site HindIII utilisé étant le site 10 se trouvant dans le multisite de clonage du pXL59; le pXL723 est décrit à la figure 11 tout comme le pXL545.

Des signes + ou - sont représentés au dessus de chacun de ces inserts pour indiquer s'il y a complémentation par le plasmide en question des insertions chromosomiques représentées au-dessus. C, ClaI; E, EcoRI; 15 H, HindIII; RV, EcoRV; Sau, Sau3AI; S, SstI.

Figure 13:

Construction des plasmides pXL253 et pXL367. Le fragment EcoRI de 8.7 kb est excisé puis purifié à partir du plasmide pXL151. Il est cloné au site EcoRI du pKT230 pour donner le pXL253. Ce même fragment est 20 inséré au site EcoRI du pRK290 (Ditta et al., 1981) pour donner le pXL367. 1, fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); ori, origine de réplication; nic, site de relaxation; mob, locus essentiel -pour la mobilisation (Bagdasarian et al., 1981); B, BamHI; C, ClaI; 25 E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XbaI; Xb, XbaI; tetr gène de résistance à la tétracycline; Kmr, gène de résistance à la kanamycine.

Figure 14:

Etudes des insertions des transposons Tn3lacZ et Tn5 sur le fragment 30 EcoRI de 8.7 kb cloné dans pRK290 (Ditta et al., 1980). Les insertions des transposons Tn3lacZ sont soulignées au contraire de celles des transposons Tn5. Les phases ouvertes déduites de la séquence (cobF à cobM) sont portées sur cette figure et les huit groupes d'insertions inactivationnelles (numérotés de 1 à 8 sont présentés; des signes + ou 35 - sont représentés sous chaque insertion de transposon pour indiquer

que l'insertion est inactivationnelle (-) ou non (+) pour la complémentation de différents mutants). Il y a une absence de complémentation lorsque le mutant recombiné synthétisé moins de trois fois moins de vitamine B12 que le niveau de synthèse de la souche dont 5 le mutant dérive. Ces groupes d'insertions inactivationnelles correspondent aux mutants suivants: 1, G615; 2, G614 et G616; 3, G613 et G164; 4, G620; 5, G638; 6, G610 et G609; 7, G612; 8, G611. Ces 10 mutants sont des mutants Cob d'*Agrobacterium tumefaciens* déjà décrits (Cameron et al, 1989). Une carte de restriction du fragment de 8.7 kb est portée au bas de la figure.

Figure 15:

Les séquences codantes de chacun des gènes du fragment de 5,4 kb, respectivement cobA à cobE, sont indiquées. La séquence des protéines COBA à COBE codées par ces séquences figurent sous leur séquence 15 codante respective, cobA à cobE. La composition en acides aminés de chaque protéine, en nombre et en pourcentage, respectivement de COBA à COBE, sont présentées ainsi que le poids moléculaire, l'index de polarité, le point isoélectrique, la densité optique à 260 nm et à 280 nm d'une solution à 1 mg/ml de protéine purifiée. Le profil 20 d'hydrophilicité de chaque protéine, respectivement de COBA à COBE, est représenté; il a été calculé d'après le programme de Hopp et Woods (1981). Les valeurs positives correspondent à des régions de la protéine qui sont hydrophiles. En abscisse la position des acides aminés est indiquée tandis qu'en ordonnée la valeur de l'index 25 d'hydrophilicité est figurée; lorsque cette valeur est positive, cela indique que la région de la protéine est hydrophile.

Figure 16:

Les séquences codantes de chacun des gènes du fragment de 8,7 kb, respectivement cobF à cobM, sont indiquées. Les séquences des 30 protéines COBF à COBM codées par ces séquences, figurent sous leur séquence. La légende est identique à celle de la figure 15. NB. Nous avons fait commencer la protéine COBF à l'ATG situé à la position 736; il est possible que l'ATG situé à la position 751 soit le véritable codon d'initiation de cette protéine.

35 Figure 17:

Réaction catalysée par l'acide cobyrinique a,c-diamide synthase. L'ACDAS catalyse l'amidation des fonctions acides carboxyliques des chaînes acétate périphériques a et c de l'acide cobyrinique (hydrogénobyrinique) pour donner l'acide cobyrinique diamide (hydrogénobyrinique diamide); le donneur du groupement amine utilisé dans le test enzymatique est la L-glutamine; il donne par désamination l'acide L-glutamique. X correspond aux ligands axiaux du cobalt qui peuvent être différents l'un de l'autre.

Figure 18: Réaction catalysée par la SP2MT. La SP2MT catalyse le transfert d'un méthyle du SAM au dihydrosirohydrochlorine ou précorrine-2 pour donner le précorrine-3. Le groupement méthyle est transféré à la position C20 du noyau porphyrine.

Figure 19: Structure de l'acide hydrogénobyrinique et de l'acide hydrogénobyrinique a,c-diamide.

Figure 20:

Affinités des protéines COBA et COBF pour le SAM. Les courbes donnent en unités arbitraires la radioactivité à la sortie de la colonne de TSK-125 pour chaque protéine appliquée sur cette colonne. Les temps de rétention sont indiqués en minutes et le pic de radioactivité correspondant au SAM libre est observé au temps de 10 min 30 sec.

Figure 21:

Comparaison des séquences de COBA et de COBI. Seules les régions 1, 2 et 3, de forte homologie, sont présentées. Les signes = sont placés entre les résidus identiques et - entre les résidus homologues (H K R, L I V M, A G S T, Y F W, D E Q N B Z, P, C).

Figure 22:

Comparaison des séquences primaires des protéines COBA de *Pseudomonas denitrificans* et CYSG de *E. coli*. L'alignement a été fait suivant le programme de Kanehisa, 1984. Les signes = sont placés entre les résidus identiques et - entre les résidus homologues (H K R, L I V M, A G S T, Y F W, D E Q N B Z, P, C). Les régions 1, 2, 3 correspondent aux zones de forte homologie entre les protéines.

Figure 23:

Comparaison des séquences de CYSG de *E. coli* avec des protéines COB de *Pseudomonas denitrificans* (COBA, COBF, COBI, COBJ, COBL et COBM). Les

comparaisons portent sur les régions 1, 2 et 3, de fortes homologies, qui existent entre CYSG, COBA et COBI. Les positions sur les séquences protéiques des régions présentant des homologies sont présentées sur la figure. Nous avons considéré les groupes de résidus homologues suivants: H K R, L I V M, A G S T, Y F W, D E Q N B Z, P, C. Si, à une même position, il se trouve au moins 3 résidus homologues, nous avons encadré ces acides aminés.

Figure 24:

Construction des plasmides pXL1148 et pXL1149. Le pXL1148 est construit comme suit: le fragment BamHI-BamHI-SstI-SstI de 1.9 kb du fragment de 8,7 kb contenant les gènes cobH et cobI est purifié, et des "linkers" XbaI et EcoRI sont placés respectivement aux extrémités BamHI et SstI. Ce fragment est ensuite inséré entre les sites XbaI et EcoRI du plasmide à large spectre d'hôte pXL59 (Cameron et al., 1989) pour donner le plasmide pXL1148. Le pXL1149 est construit comme le pXL1148 si ce n'est que le fragment initialement purifié est le fragment BamHI-BamHI-SstI de 1.5 kb au lieu du fragment contenant en plus le petit fragment SstI de 400 pb utilisé pour le pXL1148. Le fragment subit ensuite les mêmes traitements enzymatiques et le même clonage dans le pXL59. 1, fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); 3, fragment BamHI-SstI contenant l'opéron lactose de *E. coli* sans promoteur, opérateur, site d'initiation de la traduction et les 8 premiers codons non essentiels de lacZ (Casadaban et al., 1983); 4, fragment Sau3AI de *Pseudomonas putida* KT2440 (Bagdasarian et al., 1981); ori, origine de réPLICATION; nic, site de relaxation; Kmr gène de résistance à la kanamycine; mob, locus essentiel pour la mobilisation (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XbaI; Xb, XbaI.

Figure 25:

Protéines totales des souches SC510 Rifr, SC510 Rifr pKT230, SC510 Rifr pXL1148, SC510 Rifr pXL1149 analysées en PAGE-SDS à 10% comme cela est décrit. Les bactéries ont été cultivées pendant 4 jours en milieu PS4, puis des lysats des protéines totales ont été effectués. Piste 1, SC510 Rifr; piste 2, SC510 Rifr pXL1149; piste 3, SC510 Rifr

pXL1148; piste 4, SC510 Rifr pKT230. Les masses moléculaires des marqueurs de masse moléculaire sont indiquées. Les positions où les protéines COBI et COBH migrent sont indiquées.

Figure 26:

5 Construction des plasmides pXL1496 et pXL1546. Le plasmide pX1496 permet de surexprimer la protéine COBF chez E. coli et le plasmide pXL1546 permet de surexprimer COBF chez Pseudomonas denitrificans. Le fragment EcoRI-XhoI de 2.2 kb est excisé et purifié à partir du fragment de 8.7 kb. Il est cloné au site EcoRI du phage M13mp19 pour 10 donner le plasmide pXL1405. Ensuite un site NdeI est introduit par mutagénèse dirigée, comme cela est décrit précédemment, à la position 733 de ce fragment; de cette manière une site NdeI se trouve juste sur le codon d'initiation présumé du gène cobF. Le nouveau plasmide ainsi obtenu est nommé pXL1406. Un fragment NdeI-SphI-SphI de 1.5 kb, 15 contenant le gène cobF à partir de son codon d'initiation présumé, est purifié après digestion partielle par les enzymes appropriées et ligaturé avec les fragments appropriés du plasmide pXL694 (fragment EcoRI-NdeI de 120 pb contenant des signaux d'expression de E. coli -voir texte- et fragment EcoRI-SphI de 3.1 kb contenant le gène de résistance à l'ampicilline, les fonctions de réPLICATION du plasmide ainsi que les terminateurs de l'opéron rrnB de E. coli comme cela est décrit dans le texte). Le plasmide ainsi construit est nommé pXL1496. 20 Le pXL1546 est construit comme suit: le fragment EcoRI-BamHI-BamHI de 2 kb du pXL1496 est purifié par digestion partielle avec les enzymes appropriées; ce fragment contient les signaux d'expression de E. coli, suivis du gène cobF, puis la partie 5' du gène cobG, elle-même suivie des terminateurs de l'opéron rrnB de E. coli comme cela est décrit dans le texte; ce fragment est cloné dans le plasmide multihôte pKT230 (Bagdasarian et al., 198-1) décrit sur la figure 30. B, BamHI; C, 25 ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI; Km<sup>r</sup>, gène de résistance à la kanamycine; Amp, gène de résistance à l'ampicilline.

Figure 27:

Protéines totales des souches SC510 Rifr, SC510 Rifr pKT230, SC510 35 Rifr pXL1546 analysées en PAGE-SDS à 10% comme cela est décrit. Les

bactéries ont été cultivées pendant 4 jours en milieu PS4, puis des lysats des protéines totales ont été effectués. Piste 1, SC510 Rifr; piste 2, SC510 Rifr pKT230; piste 3, SC510 Rifr pXL1546. Les masses moléculaires des marqueurs de masse moléculaire sont indiquées. La 5 position où la protéine COBF migre est indiquée.

Figure 28:

Protéines totales des souches *E. coli* B et *E. coli* B pXL1496 analysées en PAGE-SDS à 10 % comme cela est décrit. Piste 1, *E. coli* pXL1496 cultivée en absence de tryptophane; piste 2, *E. coli* pXL1496 cultivée 10 dans les mêmes conditions en présence de tryptophane; piste 3, *E. coli* cultivée en absence de tryptophane; piste 4, *E. coli* cultivée dans les mêmes conditions en présence de tryptophane. Les masses moléculaires des marqueurs sont indiquées. La position de migration de la protéine COBF est indiquée.

15 Figure 29:

Construction des plasmides pXL525 et pXL368. Le pXL368 est construit comme suit: le fragment EcoRV-ClaI de 2.4 kb (contenant les gènes cobA et cobE) est purifié à partir du plasmide pXL556 (B. Cameron et al., 1989) ce qui permet d'obtenir ce fragment avec un site BamHI et un 20 site XbaI aux extrémités; ce fragment est cloné dans le pXL203 aux sites BamHI et XbaI. Pour la construction du pXL525 un "linker" XbaI est ajouté au site EcoRI situé à l'extrémité droite du fragment EcoRI de 8.7 kb; ce fragment EcoRI-XbaI de 8.7 kb est ensuite cocloné avec le fragment EcoRI-XbaI de 2.4 kb provenant du pXL556 et contenant cobA 25 et cobE. Les sites de restrictions qui sont figurés entre parenthèse correspondent à des sites qui ont disparu après traitement avec le large fragment de l'ADN polymérase I de *E. coli*. 1 , fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); ori, origine de réPLICATION; 30 nic, site de relaxation; mob , locus essentiel pour la mobilisation; Kmr, gène de résistance à la kanamycine (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XbaI; Xb, XbaI; tet, gène de résistance à la tétracycline; Ampr et Amp, gène de résistance à l'ampicilline.

35 Figure 30:

Plasmides, du groupe d'incompatibilité Q, à large spectre d'hôte chez les bactéries gram-négatives. Ces plasmides sont décrits dans une précédente publication (Cameron et al., 1989) et sont utilisés dans la présente invention. 1, fragment PstI-SstI de RSF1010 (De Graff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); 3, fragment BamHI-SstI contenant l'opéron lactose de *E. coli* sans promoteur, opérateur, site d'initiation de la traduction et les 8 premiers codons non essentiels de lacZ (Casadaban et al., 1983); 4, fragment Sau3AI de *Pseudomonas putida* KT2440 (Bagdasarian et al., 1981); ori, origine de réPLICATION; nic, site de relaxation; Kmr gène de résistance à la kanamycine; Smr, gène de résistance à la streptomycine; mob, locus essentiel pour la mobilisation (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI.

15 **Figure 31:**

Temps de rétention de différents étalons corrinoïdes (1 mg/étalon) sur le système de séparation de décrit à l'exemple 9. La colonne utilisée est une colonne de Nucléosil C-18 (Macherey-Nagel). Au niveau de chaque pic d'absorbance un numéro est représenté correspondant au corrinoïde décrit ci-dessous. Le temps de rétention est porté en abscisse et l'absorbance à 371 nm se trouve en ordonnée. 1, acide cobyrinique; 2, acide cobyrinique  $\alpha$ -amide; 3, acide cobyrinique  $\gamma$ -amide; 4, acide cobyrinique  $\alpha,\gamma$ -diamide; 5, acide cobyrinique  $c$ -amide; 6, acide cobyrinique  $c,\gamma$ -diamide; 7, acide cobyrinique  $\alpha,c$ -diamide; 8, acide cobyrinique triamide; 9, acide cobyrinique tetraamide; 10, acide cobyrinique pentaamide; 11, acide cobyrique; 12, GDP-cobinamide; 13, cobinamide phosphate; 14, cobinamide; 15, cyanocobalamine 5'-phosphate; 16, cyanocobalamine.

20 **Figure 32:**

30 Séquence nucléotidique des deux brins du fragment SalI-SalI-SalI-SalI-BglI de 4748 pb de *Pseudomonas denitrificans*. Le brin situé en haut est à lire de 5' vers 3' dans le sens gauche droite qui correspond à l'orientation gauche droite du fragment de la carte de restriction présentée sur la figure 6.

35 **Figure 33:** Séquence nucléotidique des deux brins du fragment SstI-

SstI-BamHI de 3855 pb de *Pseudomonas denitrificans*. Le brin situé en haut est à lire de 5' vers 3' dans le sens gauche droite qui correspond à l'orientation gauche droite du fragment de la carte de restriction présentée sur la figure 6.

5 **Figure 34:**

Analyse des probabilités des phases codantes d'après l'usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur les six phases de lecture du fragment SalI-SalI-SalI-SalI-BglI de 4748 pb. Pour les phases qui appartiennent au même brin codant, la 10 phase la plus probable correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase. 4a. Analyse de la séquence correspondant aux nucléotides 200 à 800. Cette analyse permet d'identifier la phase ouverte 14. Elle commence à l'ATG en position 660 et se termine au 15 TGA en position 379. 4b. Analyse de la séquence correspondant aux nucléotides 800 à 1500. Cette analyse permet d'identifier la phase ouverte 15. Elle commence au GTG en position 925 et se termine au TAA en position 1440. 4c. Analyse de la séquence correspondant aux nucléotides 1450 à 2600. Cette analyse permet d'identifier la phase 20 ouverte 16. Elle commence à l'ATG en position 1512 et se termine au TGA en position 2510. 4d. Analyse de la séquence correspondant aux nucléotides 2500 à 4650. Cette analyse permet d'identifier la phase ouverte 17. Elle commence au GTG en position 2616 et se termine au TGA en position 4511.

25 **Figure 35:**

Analyse des probabilités des phases codantes d'après l'usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur les six phases de lecture du fragment SstI-SstI-BamHI de 3855 pb. Pour les phases qui appartiennent au même brin codant, la phase la 30 plus probable correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase. 5a. Analyse de la séquence correspondant aux nucléotides 1 à 905. Cette analyse permet d'identifier la phase ouverte 18. Elle commence à l'ATG en position 809 et se termine au TGA 35 en position 108. 5b. Analyse de la séquence correspondant aux

nucléotides 955 à 2105. Cette analyse permet d'identifier la phase ouverte 19. Elle commence à l'ATG en position 1971 et se termine au TGA en position 1063. 5c. Analyse de la séquence correspondant aux nucléotides 2000 à 3300. Cette analyse permet d'identifier la phase 5 ouverte 20. Elle commence à l'ATG en position 2099 et se termine au TAG en position 3115. 5d. Analyse de la séquence correspondant aux nucléotides 3250 à 3855. Cette analyse permet d'identifier la phase ouverte 21. Elle commence à l'ATG en position 3344 et se termine au TGA en position 3757.

10 Figure 36:

Construction des plasmides pXL233, pXL843 et pXL1558 à partir du pXL154. Les plasmides sont construits de la manière suivante. Le fragment EcoRI de 3,5 kb, contenant le gène cobS tronqué et la séquence en amont, est excisé du pXL154, puis purifié et cloné dans 15 le site EcoRI du pKT230. Le plasmide ainsi construit est nommé pXL233. Le fragment EcoRI-XhoI-XhoI de 3,5 kb, contenant le gène cobT et la séquence en aval, est excisé et purifié à partir du pXL154 par digestions partielles. Le fragment EcoRI-EcoRI-EcoRI de 4,3 kb contenant le gène cobS et la séquence en amont est excisé et purifié 20 à partir du pXL154, puis ligaturé au fragment de 3,5 kb précédent. Le fragment EcoRI-XhoI de 8 kb environ ainsi obtenu est cloné dans les sites EcoRI et SalI du pXL59 pour générer le plasmide pXL843. Le plasmide pXL1558 est construit de la façon suivante: le fragment HindIII-HindIII de 12 kb est excisé du pXL154 et purifié puis les 25 extrémités sont remplies par le grand fragment de l'ADN polymérase I de *E. coli*. Cet insert est cloné dans le pRK290 (Ditta et al. 1981) digéré par EcoRI puis traité avec le large fragment de l'ADN polymérase I de *E. coli* pour rendre les extrémités franches. Les sites de restriction qui sont figurés entre parenthèse correspondent à 30 des sites qui ont disparus au cours du clonage. 1, fragment PstI-SstI de RSF1010 (Degraff et al., 1978); 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); B, BamHI; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XhoI; Xb, XbaI; Tet gène de résistance à la tétracycline; Kmr gène de résistance à la kanamycine; Smr, gène 35 de résistance à la streptomycine.

**Figure 37:**

Etude des insertions du transposon Tn5Sp sur l'insert HindIII-HindIII de 12 kb du pXL154. Les insertions du transposon sont cartographiées sur l'insert HindIII-HindIII de 12 kb, cloné sur le pXL1558. Les 5 insertions chromosomiques dans la souche SC510 Rifr sont encadrées, celle qui ne l'est pas, est introduite dans la souche SBL27 Rifr. Un signe plus ou moins est représenté sous chaque insertion pour indiquer le phénotype Cob de la souche ayant cette insertion. L'absence de complémentation (ou la complémentation) de la souche G2035 par des 10 plasmides pXL1558::Tn5Sp est indiquée par des signes moins (ou plus) au dessous de chaque insertion. Les inserts des plasmides, décrits dans la figure 36, sont représentés. Les signes plus (ou moins), sur ces plasmides et alignés avec les insertions au transposon, schématisent la complémentation (ou (l'absence) de la souche mutée au 15 transposon par le plasmide. Les phases ouvertes déduites de la séquence sont aussi portées sur cette figure (ORF14 à 17 ainsi que les gènes cob correspondants (cobS et cobT). E : EcoRI; H : HindIII; X : XhoI.

**Figure 38:**

20 Construction des plasmides pXL1286, pXL1303, pXL1324, pXL1490B et pXL1557 à partir du pXL519. La position du fragment séquencé est présentée dans la partie supérieure de la figure au dessus de la carte de restriction du cluster; il s'agit d'un SstI-SstI-BamHI de 3.9 kb. Les plasmides sont construits de la manière suivante. Le fragment 25 BglIII-EcoRI de 2 kb, contenant le gène cobU et la séquence en aval, est excisé du pXL519 puis purifié et cloné aux sites BamHI et EcoRI du pKT230 pour générer le plasmide pXL1286. Le fragment SstI-EcoRI de 2,7 kb, contenant le gène cobV tronqué, le gène cobU et la séquence en aval, est excisé du pXL519 puis purifié et cloné aux sites SstI et 30 EcoRI du pKT230 pour générer le plasmide pXL1324. Le fragment SstI-SstI de 1,6 kb, contenant le gène cobV tronqué et la séquence en amont, est excisé du pXL519 puis purifié et cloné au site SstI du pKT230 pour générer le plasmide pXL1303. Le fragment SstI-SstI-BamHI de 3.85 kb est purifié après digestion totale du pXL519 par BamHI et 35 digestion partielle par SstI. Ce fragment est ensuite cloné aux sites

BamHI et SstI du pKT230 pour générer le pXL1490B. Le plasmide pXL1557 est construit de la façon suivante: le fragment HindIII-BamHI de 9 kb est excisé du pXL519 et purifié puis les extrémités sont remplies par le grand fragment de l'ADN polymérase I de *E. coli*. Cet insert est 5 cloné dans le pRK290 (Ditta et al. 1981) digéré par EcoRI puis traité avec le large fragment de l'ADN polymérase I de *E. coli* pour rendre les extrémités franches. Les sites de restriction qui sont figurés entre parenthèse correspondent à des sites qui ont disparus au cours du clonage. 1, fragment PstI-SstI de RSF1010 (Degraff et al., 1978); 10 2, fragment PstI-BamHI de pACYC177 (Bagdasarian et al., 1981); B, BamHI; Bg, BglII; C, ClaI; E, EcoRI; H, HindIII; P, PstI; S, SstI; Sa, SalI; X, XbaI; Xb, XbaI; Tétr gène de résistance à la tétracycline; Kmr gène de résistance à la kanamycine; Smr, gène de résistance à la streptomycine.

15 15 Figure 39: Etude des insertions du transposon Tn5Sp sur l'insert HindIII-BamHI de 9 kb du pXL519. Les insertions du transposon sont cartographiées sur l'insert HindIII-BamHI de 9 kb, cloné sur le pXL1557. Les insertions chromosomiques dans la souche SC510 Rifr sont encadrées, celles qui ne le sont pas, sont introduites dans la souche 20 SBL27 Rifr. Un signe plus ou moins est représenté sous chaque insertion pour indiquer le phénotype Cob de la souche ayant cette insertion. L'absence de complémentation (ou la complémentation) de la souche G2040 par des plasmides pXL1557::Tn5Sp est indiquée par des signes moins (ou plus) au dessous de chaque insertion. Les inserts des 25 plasmides, décrits dans la figure 6, sont représentés. Les signes plus (ou moins), sur ces plasmides et alignés avec les insertions au transposon, schématisent la complémentation (ou l'absence) de la souche mutée au transposon par le plasmide. Les phases ouvertes déduites de la séquence sont aussi portées sur cette figure (ORF18 à 30 21) ainsi que les gènes cob correspondant (cobU et cobV).

Figure 40:

Séquences codantes de chacun des gènes du fragment de 4,8 kb, respectivement cobX, cobS et cobT, sont indiquées. La séquence des protéines COBX, COBS et COBT codées par ces séquences figure sous les 35 séquences codantes respectives cobX, cobS et cobT. La légende est

identique à celle de la figure 15.

Figure 41:

Séquences codantes de chacun des gènes du fragment de 3,9 kb, respectivement cobU et cobV, sont indiquées. La séquence des 5 protéines COBU et COBV codées par ces séquences figure sous les séquences codantes respectives cobU et cobV. La légende est identique à celle de la figure 15.

Figure 42:

A. Protéines totales des souches *E. coli* BL21 pLYS S pET3b, *E. coli* BL21 10 pLYS S pXL1937 analysées sur PAGE-SDS 10%. Piste 1, BL21 pLYS pET3b; piste 2, *E. coli* BL21 pLYS S pXL1937.

B. Protéines totales des souches : *E. coli* BL21, *E. coli* BL21 pXL1874 et *E. coli* BL21 pXL1875 analysées sur PAGE-SDS 10%. Piste 1, *E. coli* BL21; piste 2, *E. coli* BL21 pXL1874; piste *E. coli* BL21 pXL1875.

15 Les masses moléculaires des marqueurs sont indiquées. La bande correspondant à la protéine surexprimée est indiquée par une flèche.

Figure 43: Séquence nucléotidique des deux brins du fragment SstI-SstI-SstI-SstI-BglII-BglII de 13144 pb de Pseudomonas denitrificans. Le brin situé en haut est à lire de 5' vers 3' dans le 20 sens gauche droite qui correspond à l'orientation gauche droite du fragment de la carte de restriction présentée sur la figure 46.

Figure 44: Carte de restriction du fragment SstI-SstI-SstI-SstI-BglII-SstI-BglII de 13144 pb de Pseudomonas denitrificans. La ou les positions de sites de restriction courant sont 25 indiquées par ordre croissant du nombre de coupure sur le fragment séquencé; les positions correspondent à la séquence présentée sur la figure 43.

Figure 45: Analyse des probabilités des phases codantes d'après l'usage des codons en utilisant le programme de Staden et MacLachlan (1982) sur 30 les six phases de lecture du fragment SstI-SstI-SstI-BglII-SstI-BglII de 13144 pb de Pseudomonas denitrificans. Pour les phases qui appartiennent au même brin codant,

la phase la plus probable correspond à celle où une ligne pointillée, non interrompue par des codons stop, est placée sous la ligne de probabilité de cette phase.

1. Séquence correspondant aux nucléotides 1 à 2266. Cette analyse permet d'identifier la phase ouverte 22. Elle commence à l'ATG en position 429 et se termine au TAG en position 1884.

2. Séquence correspondant aux nucléotides 2266 à 4000. Cette analyse permet d'identifier la phase ouverte 23. Elle commence à l'ATG en position 3364 et se termine au TGA en position 3886.

3. Séquence correspondant aux nucléotides 3800 à 5000. Cette analyse permet d'identifier la phase ouverte 24. Elle commence à l'ATG en position 3892 et se termine au TAG en position 4954.

4. Séquence correspondant aux nucléotides 5000 à 9000. Cette analyse permet d'identifier la phase ouverte 25. Elle commence à l'ATG en position 5060 et se termine au TAG en position 8885.

5. Séquence correspondant aux nucléotides 9000 à 9700. Cette analyse permet d'identifier la phase ouverte 26. Elle commence à l'ATG en position 9034 et se termine au TGA en position 9676.

6. Séquence correspondant aux nucléotides 9600 à 13144. Cette analyse permet d'identifier les phases ouvertes 27, 28, 29 et 30. Elles commencent respectivement aux ATG en position 9678, 10895, 11656 et 13059 et se terminent aux codons stop en position 10101, 10304, 12181 et 12366. Les phases ouvertes 28 et 30 se trouvent sur le brin complémentaire de celui codant correspondant à toutes les autres phases ouvertes.

Figure 46: Fragment EcoRI-BglIII-EcoRI-BglII de 13.4 kb, positions des insertions des transposons Tn5Sp sur le fragment EcoRI de 9,1 kb, 30 positions des insertions des transposons Tn5 sur l'insert du plasmide pXL189 ainsi que les inserts des divers plasmides utilisés lors des expériences de complémentation des souches SC510 Rif<sup>r</sup> ::Tn5Sp. Les complémentations des mutants SC510 Rif<sup>r</sup> ::Tn5Sp par les plasmides sont indiquées (+)-entre 5 % et 100 % du niveau de la souche parentale, SC510

Rif<sup>r-</sup>, (-)-complémentation partielle, entre 0.5 à 5 % du niveau de SC510 Rif<sup>r-</sup>, ou (-)-absence de complémentation, c'est à dire, moins de mille fois moins que SC510 Rif<sup>r-</sup> positionnés juste au dessus des traits schématisant l'insert des plasmides et alignés avec les sites d'insertion des mutant correspondants. Au dessous de la cartographie des insertions des transposons Tn5 sur l'insert du plasmide pXL189 est représentée la complémentation (+), ou l'absence de complémentation (-) de ces plasmides mutants pour les mutants d'Agrobacterium tumefaciens G632 et G633. Sur la partie droite de la figure se trouve un tableau 5 représentant la complémentation des mutants G622, G623 et G630 (Cameron et al., 1989) par différents plasmides; (+)-complémentation totale, 100 % du niveau de la souche parentale, C58C9 Rif<sup>r-</sup>, (-)-complémentation partielle, entre 10 et 50 % du niveau de C58C9 Rif<sup>r-</sup>, ou (-)-absence de complémentation-.

10 Les différents plasmides dont l'insert est représenté sont construits comme suit (les fragments sont excisés soit du pXL156 soit du pXL157): pXL618 correspond au fragment EcoRI-BamHI de 2.5 kb cloné aux mêmes sites de pKT230 (Bagdasarian et al., 1981); pXL593 correspond au fragment BamHI de 3.1 kb cloné au site BamHI de 15 pKT230 (Bagdasarian et al., 1981); pXL623 correspond au fragment BamHI-XbaI de 1.9 kb cloné aux sites BamHI-SalI de pXL59 (Cameron et al., 1989); pXL1909 correspond au fragment BamHI-BamHI-BamHI de 8.4 kb cloné au site BamHI de pKT230 (Bagdasarian et al., 1981); 20 25 le pXL221 correspond au fragment EcoRI-ClaI de 1.6 kb cloné aux mêmes site de pXL59 (le site ClaI dans lequel est cloné ce fragment est le site ClaI du multisite de pXL59 (Cameron et al., 1989)); \* les pXL1908 et 1938 correspondent au même insert, fragment XbaI-BamHI-BamHI de 6.5 kb, auquel des linkers XbaI ont été ajoutés; 30 35 cet insert est cloné dans les deux orientations au site XbaI de pXL435 (Cameron et al., 1989); une flèche positionnée sur la figure indique la position du gène de résistance à la kanamycine vis à vis des extrémités de l'insert des deux plasmides; pXL208 correspond au fragment BamHI de 5.2 kb cloné au site BamHI de pKT230 (Bagdasarian et al., 1981);

pXL297 correspond au fragment EcoRI de 9.1 kb cloné au site EcoRI de pKT230 (Bagdasarian et al., 1981).

Les phases ouvertes (PO) définies par le séquençage du fragment (PO 22 à 30) sont représentées ainsi que les gènes cob correspondants; une flèche indique la polarité de la transcription.

5 E, EcoRI; B, BamHI; Bg, BglII; Cl, ClaI; Sau, Sau3AI; X, XhoI;

Figure 47: Séquences codantes de chacun des gènes du fragment de 13.4 kb, respectivement cobQ, cobP et cobW, cobN et cobO sont indiquées. La séquence des protéines COBQ, COBP, COBW, COBN et COBO codées par ces séquences figurent sous leur séquence codante respective cobQ,  
10 cobP, cobW, cobN et cobO. La légende est identique à celle de la figure 15.

figure 48 : A-Séquence NH<sub>2</sub>-terminale de la SUMT de M. ivanovii et séquence des oligonucléotides 923 946, 947; -, signifie qu'à cette position, le résidu n'a pas pu être déterminé; pour l'oligonucléotide  
15 antisens les acides aminés indiqués au-dessus de la séquence correspondent aux anticodons présentés. B-présentation de l'amplification enzymatique d'un fragment interne au gène de structure de la SUMT de M. ivanovii avec les oligonucléotides 946 et 947.

Figure 49 : Construction de la forme réplicative recombinante pG10. Le  
20 fragment de 615 pb obtenu par amplification est digéré par HindIII et EcoRI puis purifié comme cela est décrit. Ce fragment est ensuite mis à ligaturer avec la forme réplicative du phage M13mp19 digérée par les mêmes enzymes. Le clone recombinant est trouvé comme cela est décrit dans le texte.

25 Figure 50 : Autoradiographie d'un blot d'ADN génomique de M. ivanovii digéré par diverses enzymes, séparé par électrophorèse en gel d'agarose puis transféré sur membrane de Nylon comme cela est décrit précédemment. La membrane est hybridée avec la sonde pG10 comme cela est décrit précédemment. 1, HindIII-BglII; 2, KpnI-BglII; 3, EcoRI-BglII; 4,  
30 BglII-PstI. Les tailles des différents fragments qui hybrident avec la

sonde sont représentés en kb.

Figure 51 : Séquence nucléotidique des deux brins du fragment de 955 pb de *M. ivanovii*. Le brin situé en haut est à lire de 5' vers 3' dans le sens gauche droite.

- 5 Figure 52: Séquence codante du gène corA de *M. ivanovii* obtenue à partir de la séquence de 955 pb. La séquence primaire de la protéine CORA est aussi représentée. Les acides aminés sont représentés au-dessus de leur codon et le codon stop est figuré par une étoile. Les principales propriétés physiques de la protéine CORA de *M. ivanovii*, à savoir, la  
10 composition en acides aminés, en nombre et en pourcentage, le poids moléculaire, l'index de polarité, le point isoélectrique, la densité optique à 280 nm d'une solution à 1 mg/l de protéine purifiée. Le profil d'hydrophobicité de la protéine CORA de *M. ivanovii*; ce profil a été réalisé d'après le programme de Hopp et Woods (1981). Les valeurs  
15 positives correspondent à des régions de la protéine qui sont hydrophiles. Sont indiquées, la position des acides aminés en abscisse, et en ordonnée la valeur de l'index d'hydrophilicité; si cette valeur est positive, cela indique que dans cette région la protéine est hydrophile.
- 20 Figure 53 : Comparaison des séquences primaires des protéines COBA de *P. denitrificans* et CORA de *M. ivanovii*. Les protéines ont été alignées grâce au programme de Kanehisa (1984). =, acides aminés identiques; -, acides aminés homologues d'après les critères définis précédemment (voir figure 22 et 23).
- 25 Figure 54 : Construction des plasmides pXL1832 et pXL1841.  
Les légendes décrites placées sur la figure permettent de suivre les constructions.

Techniques générales de clonage, de biologie moléculaire et de biochimie.

Les méthodes classiques de biologie moléculaire telles que la centrifugation d'ADN plasmidique en gradient de chlorure de césum-bromure d'éthidium, les digestions par des enzymes de restriction, l'électrophorèse sur gel, l'électroélution des fragments d'ADN à partir de gels d'agarose, la transformation dans *E. coli*, etc, sont décrites dans la littérature (Maniatis et al., 1982, Ausubel et al., 1987).

Les enzymes de restriction ont été fournies par New-England Biolabs (Biolabs), Bethesda Research Laboratories (BRL) ou Amersham Ltd (Amersham). Les oligonucléotides "linkers" ont été fournis par Biolabs. Pour les ligatures, les fragments d'ADN sont séparés selon leur taille sur des gels d'agarose 0.7 % ou acrylamide 8 %, purifiés par électroélution, extraits au phénol, précipités à l'éthanol puis incubés dans un tampon Tris-HCl pH 7.4 50 mM, MgCl<sub>2</sub> 10 mM, DTT 10 mM, ATP 2 mM, en présence d'ADN ligase du phage T4 (Biolabs). Si nécessaire, les fragments d'ADN ayant des extrémités 5' proéminentes sont déphosphorylés par un traitement à la phosphatase alcaline d'intestin de veau (CIP, Pharmacia) à 37°C pendant 30 mn dans le tampon suivant: glycine 100 mM, MgCl<sub>2</sub> 1 mM, ZnCl<sub>2</sub> 1 mM, pH 10.5. La même technique est utilisée pour la déphosphorylation des extrémités 3' proéminentes ou franches, mais le traitement est de 15 mn à 37°C puis de 15 mn à 56 °C. L'enzyme est inactivée par chauffage du mélange réactionnel à 68°C pendant 15 mn -en présence de SDS 1% et de NaCl 100 mM suivi d'une extraction au phénol-chloroforme et d'une précipitation à l'éthanol. Le remplissage des extrémités 5' proéminentes est effectué par le fragment de Klenow de l'ADN polymérase I d'*E. coli* (Biolabs). La réaction est effectuée à température ambiante pendant 30 mn dans un tampon Tris-HCl pH 7.2 50 mM, dNTPs 0.4 mM, MgSO<sub>4</sub> 10 mM, DTT 0.1 mM, BSA 50 mg/ml. Le remplissage des extrémités 3' proéminentes est effectué en présence de l'ADN polymérase du phage T4 (Biolabs) selon les recommandations du fabricant. La digestion des extrémités proéminentes est effectuée par traitement limité à la nucléase S1 (BRL) selon les recommandations du fabricant. Des oligonucléotides "linkers" sont ajoutés sur des extrémités de fragments d'ADN comme cela est déjà décrit (Maniatis et al, 1982). La

mutagénèse *in vitro* par oligodéoxynucléotides est effectuée selon la méthode développée par Taylor et al., 1985, en utilisant le kit distribué par Amersham. Les ADN ligaturés sont utilisés pour 5 transformer la souche rendue compétente: *E. coli* MC1060 [ $\Delta$ (lacIOPZYA)X74, galU, galK, strAr, hsdR] pour les plasmides ou *E. coli* TG1[ $\Delta$ (lac proA,B), supE, thi, hsdd5/ F' traD36, proA+, B+, lacIq, lacZ $\Delta$ M15] pour les formes réplicatives de phages dérivés du bactériophage M13. Les ADN plasmidiques sont purifiés suivant la 10 technique de Birnboim et Doly, 1979. Les minipréparations d'ADN plasmidique sont faites suivant le protocole de Klein et al., 1980. Les préparations d'ADN chromosomiques de bactéries gram-négatives sont réalisées comme cela a été déjà décrit (Cameron et al., 1989). Les sondes radioactives sont préparées par translation de coupure suivant 15 la méthode déjà détaillée (Rigby et al., 1977). Les hybridations entre des séquences d'ADN ainsi que l'immobilisation des acides nucléiques sur membranes de nitrocellulose sont effectuées comme cela a été décrit (Cameron et al., 1989). Lors de clonages pour lesquels il y a une faible probabilité de trouver le clone recombinant recherché, 20 ceux-ci sont trouvé après hybridation sur filtre comme cela est déjà décrit (Maniatis et al., 1982). La séquence nucléotidique de fragments d'ADN est réalisée par la méthode de terminaison de chaînes (Sanger et al., 1977). Dans le mélange réactionnel le dGTP est remplacé par le 7-déaza-dGTP, ceci afin d'éviter les compressions de bandes lors de 25 l'électrophorèse sur gel d'acrylamide, provoquées par le pourcentage en GC élevé de l'ADN. Les milieux de culture utilisés pour la partie bactériologique ont déjà été présentés (Maniatis et al., 1982). Les cultures en milieu PS4 sont réalisées comme cela est déjà décrit (Cameron et al., 1989); les souches de *Pseudomonas denitrificans* SC510 30 Rifr et G2 Rifr sont cultivées en milieu PS4 comme suit: des Erlenmeyers de 250 ml contenant 25 ml de milieu PS4, avec si besoin l'antibiotique sélectif pour le plasmide porté par chaque souche, sont ensemencés avec une dilution au 1/100 de préculture saturée en milieu L (Miller 1972), avec, si besoin est, l'antibiotique sélectif pour le 35 plasmide porté par chaque souche; ces cultures sont incubées 6 jours à 30°C, puis les moûts sont analysés pour leur titre en cobalamines ou

bien l'activité enzymatique de certaines enzymes de la voie. Les souches d'*Agrobacterium tumefaciens*, *Pseudomonas putida* et de *Rhizobium meliloti* sont cultivées à 30°C, sauf spécifications, elles sont cultivées en milieu L. Les conjugaisons bactériennes sont réalisées comme cela est déjà décrit (Cameron et al., 1989). Les extraits de protéines totales sont réalisés comme cela est déjà décrit (Ausubel et al., 1987). Les électrophorèses (SDS-PAGE) d'analyse des protéines en gel d'acrylamide en conditions dénaturantes sont effectuées comme cela est déjà décrit (Ausubel et al., 1987).

L'appareil PhastSystem (Pharmacia) utilisant le système de tampons discontinus de Laemli (Laemli, 1970) est aussi utilisé; différents gels sont utilisés en fonction des poids moléculaires des protéines à analyser ainsi que leur pureté: PhastGel gradient 8-25 PhastGel Homogeneous 12.5 La coloration est effectuée soit au bleu de Coomassie à l'aide du PhastGel Blue R (Pharmacia), soit au nitrate d'argent en utilisant le PhastGel silver Kit (Pharmacia) en se conformant aux instructions du fabricant. Les séquences NH<sub>2</sub>-terminales des protéines sont déterminées par la technique de dégradation d'Edman en utilisant un séquenceur automatisé (Applied Biosystems modèle 407A) couplé à un appareil de CLHP pour l'identification des dérivés phénylthiohydantoïnes.

EXEMPLE 1 - Isolement de fragments d'ADN de *P.denitrificans* contenant des gènes Cob

Cet exemple décrit l'isolement de fragments d'ADN de *Pseudomonas denitrificans* portant des gènes Cob. Ces fragments ont été mis en évidence par des expériences de complémentation de mutants Cob d'*A.tumefaciens* et *P.putida* (Cameron et al., 1989).

Ces mutants Cob ont été obtenus par mutagénèse avec la N-méthyl-N'-nitro-N-nitrosoguanidine selon la technique de Miller (Miller et al., 1972), ou par insertions du transposon Tn5. De cette manière, des souches incapables de synthétiser les cobalamines ont été mises en évidence, et en particulier, le mutant cob G572 de *P.putida*, et les mutants Cob G159, G161, G164, G169, G171, G258, G609, G610, G611, G612, G613, G614, G615, G616, G620, G622, G623, G630, G632,

G633, G634, G638, G642, G643, G2034, G2035, G2037, G2038, G2039, G2040, G2041, G2042 et G2043 de A.tumefaciens.

Parallèlement, une banque d'ADN génomique de P.denitrificans est réalisée dans un vecteur mobilisable à large 5 spectre d'hôte, pXL59, par digestion, en présence d'enzymes de restriction, de 5 µg d'ADN (Cameron et al., 1989).

Par complémentation, plusieurs plasmides ont pu être isolés, permettant de complémenter les mutants cob de P.putida et d'A.tumefaciens. Parmi ceux-ci on notera plus particulièrement les 10 plasmides pXL151, pXL154, pXL156, pXL157 et pXL519.

Ces plasmides ont été isolés et des fragments d'ADN ont pu être excisés, purifiés et analysés par restriction. Ces fragments sont présentés sur les figures 6 et 44: Un fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb, un fragment EcoRI-EcoRI de 8,7 kb, un fragment 15 SalI-SalI-SalI-SalI-BglI de 4,8 kb, un fragment SstI-SstI-BamHI de 3,9 kb et un fragment EcoRI-BglII-EcoRI-BglII de 13,4 kb.

#### EXEMPLE 2 - Séquençage des fragments d'ADN isolés

Cet exemple illustre le séquençage de fragments d'ADN portant des gènes cob de *Pseudomonas denitrificans* SC510.

##### 20 2.1. Séquençage d'un fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb

Ce fragment est contenu dans le plasmide pXL157 décrit dans l'exemple 1. Après excision, des sous fragments du fragment de 5,4 kb ont été clonés dans les phages M13mp18, ou M13mp19 (Norrrander et al., 1983), ou M13tg130, ou M13tg131 (Kieny et al., 1983) dans les deux orientations. Des délétions ont été ensuite réalisées in vitro 25 par la méthode de Henikoff (1987). Ces délétions ont ensuite été séquencées avec le "primer universel" comme amorce de synthèse des réactions de terminaison de chaîne. Le recouvrement entre ces 30 différentes délétions a permis d'établir la séquence totale, sur les deux brins, du fragment de 5,4 kb (figure 7). Ce fragment comprend 5378 pb. Sur la séquence décrite à la figure 7 figurent, avant le site

ClaI, trois sites de restriction (PstI, SalI et XbaI) qui sont apparus lors du clonage du fragment en question en vue du séquençage dans des multisites de clonage. Lorsque par la suite nous nous référerons, dans la présente invention, à la séquence de ce fragment 5 ClaI-HindIII-HindIII-HindIII, ce sera à la séquence présentée à la figure 7 où les 22 premières bases ne correspondent pas à de l'ADN de Pseudomonas denitrificans (ainsi toutes les positions de site de restriction ou de début de phase ouverte se réfèrent à la séquence présentée à la figure 7).

10 2.2 Séquence nucléotidique d'un fragment EcoRI-EcoRI de  
8,7 kb

Ce fragment est porté par le pXL151 décrit dans l'exemple  
1. Le site EcoRI ainsi que les 70 pb adjacentes situées à la droite de ce fragment proviennent du pXL59 qui est le vecteur qui a servi à construire le pXL151 par clonage d'un fragment Sau3AI de Pseudomonas denitrificans SC510. Après excision, des sous fragments du fragment de 8,7 kb ont été clonés dans les phages M13mp18, ou M13mp19 (Norlander et al., 1983), ou M13tg130, ou M13tg131 dans les deux orientations (Kieny et al., 1983). Des délétions ont été ensuite réalisées in vitro par la méthode de Henikoff (1987). Ces délétions ont ensuite été séquencées avec le "primer universel" comme amorce de synthèse des réactions de terminaison de chaîne. Le recouvrement entre ces différentes délétions a permis d'établir la séquence totale, sur les deux brins, du fragment de 8,7 kb (figure 8). Ce fragment comprend 8753 pb.

2.3. Séquençage d'un fragment SalI-SalI-SalI-SalI-SalI-BglII de 4,8 kb.

30 Ce fragment est contenu dans le plasmide pXL154 décrit dans l'exemple 1. Le protocole est identique à celui utilisé dans l'exemple 2.2. La séquence totale sur les deux brins du fragment de 4,8 kb est présentée sur la figure 32. Ce fragment contient 4749 pb.

#### 2.4. Séquence nucléotidique d'un fragment

SstI-SstI-BamHI de 3,9 kb.

Ce fragment est inclus dans le plasmide pXL519 décrit dans l'exemple 1. Le protocole est identique à celui utilisé dans l'exemple 5 2.2. La séquence totale sur les deux brins du fragment de 3,9 kb est présentée sur la figure 33. Ce fragment contient 3855 pb.

#### 2.5. Séquence nucléotidique d'un fragment EcoRI-BglIII-EcoRI-BglIII de 13.4 kb.

Ce fragment est contenu dans les plasmides pXL156 et pXL157 10 décrits dans l'exemple 1. Le protocole utilisé est identique à celui de l'exemple 2.2. La séquence sur les deux brins du fragment de 13.15 kb est présentée sur la figure 43. Elle correspond à la séquence totale du fragment de 13.4 kb mis à part 250 pb, correspondant à un fragment EcoRI-SstI, se trouvant à l'extrémité gauche de celui-ci.

15 A partir de ces séquences nucléotidiques, des cartes de restriction ont été obtenues pour les enzymes qui coupent le moins fréquemment (figures 6 et 44). Le pourcentage en bases GC de l'ADN de Pseudomonas denitrificans SC150 est relativement élevé (65,5 %) et se traduit par des compressions sur les gels de séquence. Pour éviter ces 20 problèmes deux solutions sont apportées :

i) l'utilisation du 7-déaza-dGTP au lieu du dGTP dans les réactions de séquence pour diminuer les structures secondaires qui se forment pendant l'électrophorèse dans le gel de séquence et

ii) le séquençage des deux brins.

25 EXEMPLE 3 - Analyse de ces séquences nucléotidiques : détermination des phases ouvertes

Les séquences nucléotidiques des fragments ClaI-HindIII-HindIII-HindIII de 5,4 kb (figure 7), EcoRI-EcoRI de 8,7 kb (figure 8), SalI-SalI-SalI-SalI-BglI de 4,8 kb (figure 32), 30 SstI-SstI-BamHI de 3,9 kb (figure 33) et EcoRI-BglIII-EcoRI-BglII de

13,4 kb (figure 43) permettent de définir des phases ouvertes. Comme il s'agit d'ADN à haut pourcentage en GC, les phases ouvertes sont nombreuses vu la faible fréquence de codons d'arrêt de traduction. Une étude de la probabilité des phases codantes d'après usage des codons en utilisant la méthode de Staden et MacLachlan (1982) est réalisée.

5 Elle caractérise les phases ouvertes qui ont la meilleure probabilité d'être codantes par rapport aux autres phases du même brin d'ADN, cette probabilité est fonction de l'usage des codons de gènes déjà séquencés provenant des bactéries du genre Pseudomonas. De cette manière :

10 3.1. Cinq phases ouvertes sont caractérisées pour le fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb. Elles sont nommées phases 1 à 5 et leurs positions sur la séquence du fragment de 5,4 kb sont les suivantes (sur la séquence 5'-3' du site ClaI vers les sites HindIII):

15 Tableau : Phases ouvertes probables du fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb. Les positions sur la séquence correspondent aux positions sur la séquence décrite à la figure 7 ; le brin codant est le brin 5'-3' correspondant au brin supérieur sur cette figure.

	Numéro de la phase	Codon de démarrage de la traduction	Codon stop	Poids moléculaire en kD de la protéine codée
20	1	549	1011	15.5
	2	1141	1981	29.2
25	3	1980	3282	45.7
	4	3281	4280	35.0
	5	4284	5253	34.1

Les représentations des probabilités que ces phases ouvertes soient des phases codantes, avec parallèlement celles

observées sur les autres phases (5 au total) sont portées sur la figure 9. Ces cinq phases sont codées par le même brin. Quatre d'entre elles (phases ouvertes 1 à 4) présentent les caractéristiques de phases codantes en couplage traductionnel (Normak et al., 1983), à savoir que le codon d'initiation de traduction de la phase  $x+1$  chevauche le codon de terminaison traduction de la phase  $x$  ou bien que ceux-ci sont très proches.

3.2. Huit phases sont caractérisées pour le fragment EcoRI-EcoRI de 8,7 kb. Elles sont nommées phases 6 à 13 et leurs positions sur la séquence du fragment de 8,7 kb sont portées sur le tableau ci-dessous.

Tableau : Phases ouvertes probables du fragment EcoRI 8,7 kb. Les positions sur la séquence correspondent aux positions sur la séquence décrite sur la figure 8 ; sur cette figure, le brin codant est le brin supérieur sauf pour la phase 11.

	Numéro de la phase de traduction	Codon de démarrage	Codon stop	Poids moléculaire en kD de la protéine codée
20	6	736	1519	28.9
	7	1620	2997	46.7
	8	3002	3632	22.0
	9	3631	4366	25.8
	10	4365	5127	27.1
	11	5893	5110	28.0
	12	5862	7101	42.9
25	13	7172	7931	26.8

Les représentations des probabilités de ces phases ouvertes, avec parallèlement celles observées sur les autres phases (6 phases au total) sont portées sur la figure 10. A l'exception de la

phase 11, ces phases sont codées par le même brin. Quatre d'entre elles (de 7 à 10) présentent les caractéristiques de phases codantes en couplage traductionnel (Normark et al., 1983), à savoir que le codon d'initiation de traduction de la phase  $x+1$  chevauche le codon de terminaison de traduction de la phase  $x$  ou bien que ceux-ci sont très proches.

10 3.3. Quatre phases ouvertes sont caractérisées pour le fragment SalI-SalI-SalI-SalI-BglII de 4,8 kb. Elles sont nommées phases 14 à 17 et leurs positions sur la séquence du fragment de 4,8 kb sont les suivantes (sur la séquence 5'- 3' des sites SalI vers le site BglII):

15 Tableau: Phases ouvertes probables du fragment SalI-SalI-SalI-SalI-BglII de 4,8 kb. Les positions sur la séquence correspondent aux positions décrites sur la figure 32 où le brin supérieur est donné dans son orientation 5'- 3'. Les phases 15, 16 et 17 sont codées par le brin supérieur au contraire de la phase 14

	Numéro de la phase	Codon de démarrage de traduction	Codon stop	Poids moléculaire en kD de la protéine codée
20	14	660	379	10286
	15	925	1440	18941
	16	1512	2510	36983
	17	2616	4511	70335

25 Les représentations des probabilités que ces phases ouvertes soient codantes, avec parallèlement celles observées sur les autres phases (au total 4) sont portées sur la figure 34. Les phases 15, 16 et 17 sont codées par le même brin, la phase 14 par le brin complémentaire.

3.4. Quatre phases sont caractérisées pour le fragment SstI-SstI-BamHI de 3,9 kb. Elles sont nommées 18 à 21 et leurs positions sur la séquence du fragment de 3,9 kb sont portées sur le tableau ci-dessous.

5 **Tableau:** Phases ouvertes probables du fragment SstI-SstI-BamHI de 3,9 kb. Les positions sur la séquence correspondent aux positions décrites sur la figure 33 où la polarité du brin supérieur est 5'- 3'. Les phases 18 et 19 sont codées par le brin inférieur au contraire des phases 20 et 21.

	Numéro de la phase	Codon de démarrage de traduction	Codon stop	Poids moléculaire en kD de la protéine codée
10	18	809	108	25148
	19	1971	1063	30662
	20	2099	3115	34682
	21	3344	3757	14802

20 Les représentations des probabilités que ces phases ouvertes soient codantes, avec parallèlement celles observées sur les autres phases (au total 4) sont portées sur la figure 35. Les phases 19 et 20 sont transcrives de façon divergente.

3.5. Neuf phases ouvertes sont caractérisées pour le fragment EcoRI-BglII-EcoRI-BglII de 13,1 kb. Elles sont nommées phases 22 à 30 et leurs positions sur la séquence du fragment de 13,1 kb sont les suivantes (sur la séquence 5'- 3' du site EcoRI vers le site BglII):

Tableau: Phases ouvertes probables du fragment EcoRI-BglII-EcoRI-BglII de 13,1 kb. Les positions sur la séquence correspondent aux positions

décrivées sur la figure 43 où le brin supérieur est donné dans son orientation 5'- 3'. Les phases 22, 23, 24, 25, 26, 27 et 29 sont codées par le brin supérieur au contraire des phases 28 et 30.

	Numéro de la phase de traduction	Codon de démarrage	Codon stop	Poids moléculaire en kD de la protéine codée
5	22	429	1884	51 982
	23	3364	3886	19442
	24	3892	4954	38121
	25	5060	8885	138 055
	26	9034	9676	24 027
	27	9678	10101	14 990
	28	10835	10306	21 057
	29	11656	12181	19 183
15	30	13059	12368	24 321

Les représentations des probabilités que les phases ouvertes 22, 23, 24, 25, et 26 soient codantes, avec parallèlement celles observées sur les autres phases (au total 5 phases) sont portées sur la figure 45.

20 Ces 5 phases sont codées par le même brin.

EXEMPLE 4 - Etudes génétiques sur les fragments d'ADN portant des gènes cob

Cet exemple montre la relation qui existe entre les différentes phases ouvertes identifiées précédemment et les gènes impliqués dans la biosynthèse des cobalamines et/ou des cobamides portés par ces mêmes fragments. Ces gènes sont identifiés par une étude génétique comme décrit ci-dessous.

4.1 - Etude génétique du fragment de 5,4 kb

Le plasmide pXL723 est le plasmide pRK290 (Ditta et al., 1980) contenant le fragment EcoRI-HindIII de 2264 pb, correspondant à la partie droite du fragment étudié, cloné au site EcoRI du pRK290 (figure 11). La construction des autres plasmides utilisés dans cette 5 étude (pXL302, pXL1397, pXL545, pXL545Ω, pXL556 et pXL1500) est décrite dans la légende des figures 11 et 12.

Des insertions ont été obtenues sur le plasmide pXL723 en utilisant la technique de Bruijn et Lupski, 1984. Les insertions du transposon Tn5 sur le plasmide pXL723 ont été sélectionnées, puis 10 cartographiées dans le fragment de 5,4 kb (figure 12). Le pXL723 complémente le mutant Cob G572 de Pseudomonas putida et le mutant Cob G634 d'Agrobacterium tumefaciens. Ces insertions se classent en deux groupes d'insertions inactivationnelles: soit celles qui ne permettent plus de complémenter le mutant Cob G572, soit celles qui abolissent la 15 complémentation du mutant Cob G634 (figure 12). Les insertions qui inactivent la complémentation du mutant G572 sont cartographiées dans la phase ouverte 4 (il s'agit des insertions 15, 27, 68, 81 et 97) ; la phase ouverte 4 correspond donc à un gène cob. Celui-ci est nommé cobC. Les insertions qui inactivent la complémentation du mutant G634 20 sont cartographiées dans la phase 5 (ce sont les insertions 66 et 107, figure 12); la phase ouverte 5 correspond donc à un gène cob. Celui-ci est nommé cobD. Par ailleurs des insertions avec un transposon Tn5Sp<sup>r</sup> ont été réalisées. Le transposon Tn5Sp<sup>r</sup> a été construit au laboratoire en clonant au site BamHI du transposon Tn5 (Jorgensen et al., 1979) 25 une cassette BamHI contenant le gène de résistance à la spectinomycine provenant du plasmide pHG45Ω (Prentki et Krisch, 1984). Ces insertions ont été faites dans le chromosome de la souche de Pseunomonas denitrificans SBL27 Rif<sup>r</sup>. La souche SBL27 est une souche de Pseudomonas denitrificans dont SC510 dérive par plusieurs mutagénèses. 30 SBL27 produit 10 fois moins de cobalamines que SC510 sur milieu PS4. Sur 10000 clones de la souche SBL27 Rif<sup>r</sup>, portant chacun une insertion du transposon, plus de 30 d'entre eux avaient perdu la capacité de synthétiser des cobalamines. Certains de ces clones possédaient une insertion dans le fragment étudié dans cet exemple. Ces insertions ont

étété cartographiées par analyses de restriction selon la méthode de Southern (Southern, 1975). Les sites d'insertions du transposon chez ces différents mutants sont portés sur la figure 12. Une de ces insertions numérotée 2639 se trouve dans le gène cobC; cette insertion 5 est complémentée par le plasmide pXL302 qui porte un fragment contenant le gène cobC (figure 12). Deux insertions, nommées 2636 et 2638, sont dans la phase ouverte 3. Ces mutants sont bloqués dans la biosynthèse des cobalamines, et ils sont complémentés par le plasmide pXL1397 qui ne contient que la phase ouverte 3, mais non complémentés 10 par le plasmide pXL302 qui contient les gènes cobC et cobD (figure 12). Ces deux insertions sont donc dans un autre gène. A la phase ouverte 3 nous associons le gène cobB. Une insertion 2933 est placée dans la phase ouverte 2 ; elle est complémentée par le plasmide pXL1500 qui contient la phase ouverte 2 ; cette insertion est non 15 complémentée par le plasmide pXL1397 qui contient le gène cobB et qui complémente les deux insertions dans cobB. Il s'agit donc ici d'une insertion dans un autre gène; à la phase ouverte 2 nous associons un gène nommé cobA.

Une cassette de résistance à la kanamycine provenant du 20 plasmide pUC4K (Barany et al., 1985) a été introduite au site NotI du fragment ClaI (position 0 sur la séquence)-RsaI (position 1686 sur la séquence) cloné dans un plasmide pUC8 (Viera et Messing, 1982) ; il s'agit du site NotI situé à la position 771 dans la phase 1 (voir la séquence sur la figure 7); deux insertions ont été retenues 25 correspondant chacune à une orientation différente de la cassette de résistance. Ces fragments portant chacun une insertion de la cassette de résistance ont été clonés dans le plasmide pRK404 (Ditta et al.) pour donner les plasmides pXL1630 et 1631. Ces plasmides ont été introduits par transfert conjugalif dans la souche de Pseudomonas 30 denitrificans SC510 Rif<sup>r</sup>, puis par une série de cultures-dilutions en absence de l'antibiotique sélectif pour le plasmide (la tétracycline), des doubles recombinants ayant échangé le fragment plasmidique avec le fragment chromosomique et ayant perdu le plasmide ont été trouvés. Deux souches ont ainsi été caractérisées :

35 i) l'une est nommée SC510:1631 Rif<sup>r</sup>, dans cette souche, la

cassette de résistance à la kanamycine se trouve insérée, dans le chromosome au site NotI (se trouvant dans la phase 1); les polarités des transcriptions du gène de résistance à la kanamycine et celle de la phase ouverte 1 sont opposées,

5 ii) l'autre insertion est nommée SC510:1630 Rif<sup>r</sup>; la cassette de résistance est insérée au même site, mais la transcription du gène de résistance à la même polarité que celle de phase ouverte 1 entière.

10 Ces deux souches ont toutes les deux un taux de synthèse de cobalamines au moins 100 fois inférieur à celui de SC510.

Le plasmide pXL545Ω correspond au plasmide pXL545 dans lequel la cassette de résistance à la spectinomycine, du plasmide pHP45Ω a été insérée au niveau du site BamHI. Ce plasmide (figure 12) qui contient le fragment ClaI-HindIII de 814 pb (où seule la phase ouverte 1 est entière) complémente seulement le mutant SC510:1630 Rif<sup>r</sup>. Ceci suffit à définir un nouveau gène puisque ce mutant est complémenté par un plasmide qui contient uniquement la phase ouverte 1 entière. La phase ouverte 1 correspond à un gène de la voie de biosynthèse des cobalamines et/ou des cobamides. Ce gène est nommé cobE. L'absence de 15 complémentation du mutant SC510:1631 Rif<sup>r</sup> par le plasmide pXL545Ω est peut être due au fait que les gènes cobA, cobB, cobC, cobD et cobE, ou une partie d'entre eux, appartiennent au même opéron et que 20 l'insertion dans cobE qui conserve une transcription dans le sens de transcription de l'opéron peut être complémentée seulement par 25 l'expression en trans du gène cobE. Au contraire le mutant SC510:1631 Rif<sup>r</sup> lui ne peut être complémenté que par un plasmide qui permet 30 l'expression en trans des gènes cobA à cobE.

Le fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb contient donc cinq gènes cob nommés cobA, cobB, cobC, cobD et cobE.

#### 30 4.2 - Etudes génétiques du fragment de 8,7 kb

Le plasmide pXL367 est le pRK290 (Ditta et al., 1980) contenant le fragment EcoRI de 8,7 kb cloné au site EcoRI (figure 13).

Des insertions du transposon Tn5 sur le plasmide pXL367

ont été sélectionnées en utilisant la technique déjà décrite (de Bruijn et Lupski, 1984). Les insertions dans le fragment de 8,7 kb ont été cartographiées. De la même manière des insertions du transposon Tn3lacZ ont été obtenues selon la méthode déjà décrite (Stachel et al., 1985) et cartographiées. 29 insertions du transposon Tn5 et 13 insertions du transposon Tn3lacZ ont ainsi été cartographiées. La position précise de ces insertions sur le fragment de 8,7 kb est portée à la figure 14. Des plasmides, portant chacun une seule insertion dans le fragment de 8,7 kb, ont été introduits par transferts conjugatifs dans les mutants Cob d'Agrobacterium tumefaciens G164, G609, G610, G611, G612, G613, G614, G615, G616, G620, G638. Ces mutants sont tous complémentés par le pXL367. Les insertions qui ne permettent plus la complémentation des différents mutants ont été recherchées. Elles correspondent à une insertion dans le gène responsable de la complémentation du mutant correspondant. Les résultats des complémentations des différents mutants pour leur caractère de production de cobalamines (phénotype Cob) sont portés sur la figure 14. Si le mutant recombiné produit moins de trois fois moins de cobalamines que ne produit le même mutant avec le plasmide pXL367, il est considéré comme étant non complémenté. Sur les mutants étudiés, G164, G609, G610, G611, G612, G613, G614, G615, G616, G620, G638, on observe huit classes différentes d'insertions inactivationnelles de transposons conduisant à un phénotype muté. Ces classes caractérisent des insertions par l'absence de complémentation d'un ou plusieurs mutants par les plasmides pXL367 portant ces mêmes insertions. Chaque classe correspond donc à un gène muté. On observe que les insertions appartenant à une même classe sont positionnées les unes à côté des autres. Huit classes d'insertions sont ainsi observées qui permettent de définir huit gènes. Chaque classe d'insertions définit un fragment minimum qui doit être contenu dans le gène correspondant. La figure 14 démontre un parfait recouvrement entre les régions délimitées par chaque classe, au niveau de la carte de restriction, et les phases ouvertes décrites ci-dessus (exemple 3). On constate en effet que, pour chaque classe d'insertions, les transposons sont toujours insérés dans une partie du fragment de 8,7 kb qui est contenue dans une seule

phase ouverte. A chaque classe d'insertions, est donc associée une phase ouverte et une seule. Les phases ouvertes indiquées ci-dessus codent donc chacune pour une protéine impliquée dans la voie de biosynthèse des cobalamines et/ou des cobamides. Les phases ouvertes 5 correspondent chacune à des gènes impliqués dans la biosynthèse des cobalamines et/ou des cobamides. Ces phases ouvertes sont appelées cobF, cobG, cobH, cobI, cobJ, cobK, cobL et cobM pour les phases 6 à 13 respectivement. La position de ces gènes par rapport à la carte de restriction est représentée sur la figure 14.

10

#### 4.3 - Etude génétique du fragment de 4,8 kb.

Le plasmide pXL1558 est le plasmide pRK290 (Ditta et al., 1980) contenant le fragment HindIII-HindIII de 12 kb du pXL154 (Cameron et al., 1989) cloné au site EcoRI du pRK290 (figure 36). La construction des autres plasmides utilisés dans cette étude (pXL233 et 15 pXL843) est décrite dans la légende de la figure 36.

Des insertions Tn5Sp ont été obtenues sur le plasmide pXL1558. Tout d'abord une souche contenant un transposon Tn5Sp a été construite; ceci a été fait en transformant la souche C2110 (Stachel et al., 1985) à l'aide du plasmide pRK2013Tn5Sp (Blanche et al., 1989); le plasmide 20 pRK2013Tn5Sp ayant une origine de réPLICATION ColE1 ne se réplique pas dans la souche C2110 qui est polA-. Les colonies obtenues après transformation qui sont résistantes à la spectinomycine ont donc le transposon Tn5Sp dans leur chromosome; une colonie est ensuite réisolée, puis l'insertion du transposon est ensuite transduite à 25 l'aide du phage P1 chez la souche MC1060 comme cela est décrit précédemment (Miller, 1972). La souche MC1060 Tn5Sp est transformée par le plasmide pXL1558; le plasmide pXL1558 est ensuite mobilisé par conjugaison à l'aide du pRK2013 chez C600 Rifr. Des conjugants résistants à la tétracycline (pour le plasmide pXL1558) et à la spectinomycine (pour le transposon) sont ensuite sélectionnés; de tels conjuguants doivent contenir le plasmide pXL1558 dans lequel le transposon Tn5Sp s'est inséré. Des insertions portées sur le plasmide 30 pXL1558, et plus précisément dans le fragment de 12 kb sont ensuite

cartographiées par digestion de restriction; 23 insertions sont ainsi obtenues et cartographiées sur le fragment de 12 kb; la position de ces différentes insertions sur le fragment est présentée sur la figure 37. Ces 23 insertions ont été introduites sur le chromosome de la souche SC510 Rifr après transfert conjugal du p-XL1558::Tn5Sp, puis introduction du plasmide pR751. Le plasmide pR751 est un plasmide, résistant à la triméthoprime, du même groupe d'incompatibilité que le pXL1558 (incP, Thomas et Smith, 1987). Par culture non sélective pour le pXL1558 (absence de tétracycline) mais sélective pour le pR751 et le transposon (présence de triméthoprime et de spectinomycine) l'échange de la mutation portée par le pXL1558::Tn5Sp avec le chromosome ainsi que la ségrégation du pXL1558 sont obtenus par la technique d'échange de marqueur par double recombinaison homologue comme cela est déjà décrit (Schell et al., 1988). Les souches ainsi sélectionnées portent le transposon dans leur chromosome. La double recombinaison homologue est vérifiée par la méthode de Southern (Southern, 1975). De cette manière 23 souches SC510 Rifr::Tn5Sp dans le fragment de 12 kb ont été identifiées.

D'autre part une autre insertion Tn5Sp obtenue par mutagénèse au hasard du transposon Tn5Sp dans la souche SBL27 Rifr (Blanche et al., 1989) a été cartographiée sur le fragment de 12 kb par analyse de restriction selon la méthode de Southern (Southern, 1975), voir figure 37; cette souche est nommée SBL27 Rifr::Tn5Sp 1480.

Le taux de synthèse de cobalamines est déterminé pour ces 24 souches cultivées en milieu PS4 selon le protocole déjà décrit (Cameron et al., 1989) et le phénotype Cob- est attribué aux souches produisant au moins 1000 fois moins (resp. 100) de vitamine B12 que la souche parente SC510 Rifr (resp. SBL27 Rifr), figure 37. Il est ainsi observé que 6 de ces insertions chromosomiques conduisent à un phénotype Cob- chez *P. denitrificans*; il s'agit des insertions 31.1, 41.3, 45, 55, 22.1 et 1480.

Trois plasmides pXL233, pXL837 (Cameron et al.) et pXL843 sont introduits par transferts conjugatifs dans trois souches présentant le phénotype Cob- soient SC510 Rifr::Tn5Sp 31.1, SC510 Rifr::Tn5Sp 45, SBL27 Rifr::Tn5Sp 1480. Ces trois mutants ont chacun un profil de

complémentation différent pour la synthèse des cobalamines. En effet SBL27 Rifr::Tn5Sp 1480 est complémenté par le pXL837 et le pXL843 mais pas par le pXL233; le mutant SC510 Rifr::Tn5Sp 45 n'est complémenté que par le pXL843; le mutant SC510 Rifr::Tn5Sp 31.1 est complémenté par le plasmide pXL843 ainsi que par le plasmide pXL233 (voir figure 37). Les données présentées permettent donc, d'après les résultats sur les complémentations des trois mutants, de conclure que les trois mutants sont différents et que pour chacun d'entre eux, le transposon Tn5Sp s'est inséré dans un gène cob différent.

D'autre part les plasmides pXL1558::Tn5Sp 41.3, pXL1558::Tn5Sp 45, et pXL1558::Tn5Sp 22.1 sont introduits par transferts conjugatifs dans la souche G2035 (Cameron et al., 1989) et ne la complémentent pas. Le plasmide pXL1558 complémente ce mutant au contraire du plasmide pXL1558::Tn5Sp 31.1.

Les données de phénotype et complémentation nous permettent de définir 3 classes d'insertions; chacune de ces classes est représentée par les insertions suivantes : 31.1, classe 1; 45, 41.3, 55 et 22.1, classe 2; 1480 classe 3.

Pour chaque classe d'insertions, les transposons sont toujours insérés dans une partie du fragment de 4,8 kb qui est contenue dans une seule phase ouverte (ORF14, ORF16 et ORF17, comme définies à l'exemple 3). A chaque classe d'insertions est associée une seule phase ouverte. Les phases ouvertes indiquées ci-dessus codent donc pour une protéine impliquée dans la voie de biosynthèse des cobalamines et/ou des cobinamides. Ces phases ouvertes sont appelées cobX, cobS et cobT pour les phases 14, 16 et 17. La position de ces gènes par rapport à la carte de restriction est représentée sur la figure 37. La phase ouverte 15 n'est pas un gène impliqué dans la biosynthèse du coenzyme B12.

30

#### 4.4 - Etudes génétiques du fragment de 3,9 kb

Le plasmide pXL1557 est le plasmide pRK290 (Ditta et al., 1980) contenant le fragment HindIII-BamHI de 9 kb du pXL519 cloné au site EcoRI du pRK290 (figure 38). La construction des autres

plasmides utilisés dans cette étude (pXL1286, pXL1303, pXL1324) est décrite dans la légende de la figure 38. Par ailleurs, le fragment BglII-XhoI de 2 kb (positions sur la séquence présentée figure 33 251 et 2234) du plasmide pXL519 est cloné aux sites BamHI-SalI du plasmide 5 pXL435 (Cameron et al) pour générer le plasmide pXL699.

Des insertions Tn5Sp ont été obtenues sur le plasmide pXL1557 selon la technique décrite dans l'exemple 4.3. Les insertions du transposon Tn5Sp sur le plasmide pXL1557, alors nommées pXL1557::Tn5Sp, ont été sélectionnées. Celles qui sont cartographiées dans le fragment de 9 kb 10 (figure 39) ont été introduites sur le chromosome de la souche SC510 Rifr après transfert conjugalif du pXL1557::Tn5Sp et échange de marqueur par double recombinaison homologue comme cela est décrit en 4.3.

La double recombinaison homologue est vérifiée par la méthode de 15 Southern (Southern, 1975). De cette manière 20 souches SC510 Rifr::Tn5Sp ont été identifiées.

D'autre part deux autres insertions Tn5Sp obtenues par mutagénèse au hasard du transposon Tn5Sp dans la souche SBL27 Rifr (Blanche et al., 1989) ont été cartographiées sur le fragment de 9 kb par analyse de 20 restriction selon la méthode de Southern (Southern, 1975), voir sur la figure 39 les insertions 1003 et 1147.

Le taux de synthèse de cobalamines est déterminé pour ces 22 souches cultivées en milieu PS4 selon le protocole déjà décrit (Cameron et al., 1989) et le phénotype Cob- est attribué aux souches produisant 25 1000 fois moins (resp. 100) de vitamine B12 que la souche parente SC510 Rifr (resp. SBL27 Rifr), figure 39. Seules les 4 insertions 1, 1003, 23 et 1147 se traduisent par un phénotype Cob- chez *P. denitrificans*.

Quatre plasmides pXL699, pXL1286, pXL1303 et pXL1324 sont introduits 30 par transferts conjugatifs dans les quatre souches présentant le phénotype Cob- soient SC510 Rifr::Tn5Sp 1, SBL27 Rifr::Tn5Sp 1003, SC510 Rifr::Tn5Sp 23 et SBL27 Rifr::Tn5Sp 1147. Le plasmide pXL699 complémente les deux premiers mutants (SC510 Rifr::Tn5Sp 1, SBL27 35 Rifr::Tn5Sp 1003) mais le plasmide pXL1303 ne les complémente pas; le plasmide pXL1324 complémente les deux autres mutants (SC510

Rifr::Tn5Sp 23 et SBL27 Rifr::Tn5Sp 1147) mais le plasmide pXL1286 ne les complémente pas.

D'autre part le plasmide pXL1557::Tn5Sp 1, est introduit par transfert conjuguatif dans la souche G2040 et ne la complémente pas alors que 5 les plasmides pXL1557, pXL1557::Tn5Sp 6A, pXL1557::Tn5Sp 54, pXL1557::Tn5Sp 48, pXL1557::Tn5Sp 21, pXL1557::Tn5Sp 8, pXL1557::Tn5Sp 23 aussi introduits par transferts conjuguatifs, la complémentent (voir figure 39).

Les données de phénotype et complémentation permettent de 10 définir 2 classes d'insertions. Pour chaque classe d'insertions, les transposons sont toujours insérés dans une partie du fragment de 3,9 kb qui est contenue dans une seule phase ouverte (ORF 19 et ORF20 comme définie à l'exemple 3).

A chaque classe d'insertions est associée une seule phase ouverte. Les 15 phases ouvertes indiquées ci-dessus codent pour une protéine impliquée dans la voie de biosynthèse des cobalamines et/ou des cobinamides. Ces phases ouvertes sont appelées cobV et cobU pour les phases 19 et 20. Les phases 18 et 21 ne sont pas des gènes impliqués dans la voie de biosynthèse du coenzyme B12. La position de ces gènes par rapport à la 20 carte de restriction est représentée sur la figure 39. Les insertions 48, 21 et 8 sont cartographiées entre les gènes cobU et cobV

#### 4.5 - Etudes génétiques du fragment de 13,4 kb.

##### 4.5.1. Etudes sur le fragment EcoRI-BglII de 4327 pb.

Le plasmide pXL189 (Cameron et al., 1989), qui contient au moins un 25 gène cob porte un insert de 3.1 kb qui excepté 300 pb correspond à un fragment EcoRI-ClaI de 4,26 kb (voir figure 45). Le pXL189 a été soumis à une mutagénèse au transposon Tn5, comme cela est décrit précédemment (De Bruijn et Lupski (1984)). 13 insertions ont ainsi été cartographiées dans l'insert du pXL189 comme cela est présenté sur la figure 46. Ces 30 13 plasmides mutants ainsi que pXL189 ont été conjugués chez deux mutants d'A. tumefaciens, G632 et G633, qui sont des mutants complémentés par pXL189 (Cameron et al., 1989). Seule l'insertion 58 s'est avérée être une insertion inactivationnelle. Ce résultat montre

que les deux mutants G632 et G633 correspondent à une mutation dans un même gène et que d'autre part le seul gène de P. denitrificans qui puisse être responsable de leur complémentation correspond à la phase ouverte 26 (voir figure 46) puisque l'insertion 58 est cartographiée dans cette phase ouverte; de plus il s'agit de la seule insertion parmi les 13, qui est cartographiée dans cette phase ouverte. Un gène cob, nommé cob0 est donc associé à la phase ouverte 26.

Pour savoir si les quatre autres phases ouvertes (phases ouverte 27 à 30) identifiées sur ce fragment correspondent à des gènes cob, une cassette de résistance à la spectinomycine, du plasmide pHP45Ω (Prentki et Krisch, 1984) a été spécifiquement insérée dans chacun de ces gènes puis introduite dans le chromosome de P. denitrificans SC510 Rif<sup>r</sup> par recombinaison homologue de manière à obtenir des mutants d'insertions dans chacune de ces phases ouvertes. Pour ce faire, le fragment 15 EcoRI-ClaI (positions respectives 8818 et 13082 sur la séquence présentée sur la figure 43) a été utilisé. Ce fragment, qui porte les phases ouvertes 27 à 30, a été purifié à partir du pXL157 (Cameron et al., 1989); un linker EcoRI a été ajouté sur l'extrémité ClaI après que celle-ci ait été remplie par le fragment Klenow de l'ADN polymérase 20 d'E. coli. Ce fragment a ensuite été cloné dans le plasmide pUC13 (Viera et al., 1982) au site EcoRI. Le plasmide ainsi construit a été appelé pXL332. Des insertions de la cassette de résistance à la spectinomycine du plasmide pHP45n (Prentki et Krisch, 1984) ont été réalisées sur le pXL332. Ces insertions ont été faites séparément aux 25 sites SmaI (position 9868, phase ouverte 27), BamHI (position 10664, phase ouverte 28), ClaI (position 11687, phase ouverte 29) et NcoI (position 12474, phase ouverte 30) par digestions totales ou partielles du pXL332 avec les enzymes correspondantes puis, si cela est nécessaire, remplissage de ces extrémités avec le fragment de Klenow de 30 l'ADN polymérase d'E. coli, puis ligature avec le fragment de SmaI 2 kb du pHP45n (Prentki et Krisch, 1984) contenant un gène de résistance à la spectinomycine; ces insertions sont nommées respectivement n2, n1, n3 et n4 comme cela est présenté sur la figure 46. Les fragments EcoRI portant ces différentes insertions ont ensuite été clonés sur le pRK404 35 (Ditta et al., 1985) à l'un des deux sites EcoRI. Les 4 plasmides

portant ces différentes insertions ont ensuite été introduits par conjugaison chez SC510 Rif<sup>r</sup>, comme cela est décrit précédemment. Le plasmide pR751 (Thomas et Smith, 1987) a ensuite été introduit dans les transconjuguants. L'échange des mutations portées par les 4 différents 5 dérivés de pRK404 et le chromosome de SC510 Rif<sup>r</sup> a pu être sélectionné comme cela est décrit (voir exemple 4.3). 4 souches ont ainsi été obtenues. Ces souches portent chacune une insertion de la cassette de résistance dans une des quatre phases ouvertes 27 à 30. Ces insertions ont été vérifiées par analyse de l'ADN génomique par Southern blot 10 (Southern, 1975). La production de cobalamines de ces différentes souches a été étudiée. Elles ont toutes montré un phénotype Cob+ par culture en milieu PS4. Ce résultat indique que ces phases ouvertes n'interviennent pas dans la biosynthèse du coenzyme B12. Toutefois il est possible qu'une ou plusieurs de ces phases codent pour des 15 protéines qui interviennent par exemple dans la transformation du coenzyme B12 en méthylcobalamine par exemple, c'est à dire la synthèse d'une autre cobalamine voir d'un autre corrinoïde.

#### 4.5.2 Etude du fragment EcoRI-EcoRI de 9.1 kb.

Différents plasmides sont utilisés dans cette étude; 20 le plasmide pXL1560 est le plasmide pRK290 (Ditta et al., 1980) contenant le fragment EcoRI-EcoRI de 9,1 kb du pXL156 (exemple 1) cloné au site EcoRI du pRK290 (voir figure 46). La construction des autres plasmides utilisés dans cette étude (pXL618, pXL593, pXL623, pXL1909, pXL1938, pXL1908, pXL221, pXL208, pXL297) est décrite dans la 25 légende de la figure 45.

Des insertions Tn5Sp ont été obtenues sur le plasmide pXL1560. La souche MC1060 Tn5Sp transformée par le plasmide pXL1560 a servi à obtenir des insertions du transposon Tn5Sp dans le fragment pXL1560; 27 insertions ont ainsi été obtenues et cartographiées sur le fragment de 30 9,1 kb; la position de ces différentes insertions sur le fragment est présentée sur la figure 4. Ces 27 insertions ont été introduites sur le chromosome de la souche SC510 Rif<sup>r</sup> après transfert conjugal des

pXL1560::Tn5Sp, puis introduction du plasmide pR751. Le plasmide pR751 est un plasmide, résistant à la triméthoprime, du même groupe d'incompatibilité que le pXL1560 (incP, Thomas et Smith, 1987). Par culture non sélective pour le pXL1560 (absence de tétracycline) mais 5 sélective pour le pR751 et le transposon (présence de triméthoprime et de spectinomycine) l'échange de la mutation portée par le pXL1560::Tn5Sp avec le chromosome ainsi que la ségrégation du pXL1560 sont obtenus; cette technique d'échange de marqueur par double 10 recombinaison homologue est équivalente à celle déjà décrite par Schell et al., 1988. Les souches ainsi sélectionnées portent le transposon dans leur chromosome.

La double recombinaison homologue est vérifiée par la méthode de Southern (Southern, 1975). De cette manière 27 souches SC510 Rif<sup>r</sup>::Tn5Sp possédant chacune une insertion différente du transposon 15 Tn5Sp dans le fragment de 9.1 kb ont été identifiées.

Le taux de synthèse de cobalamines est déterminé pour ces 27 souches cultivées en milieu PS4, et le phénotype Cob<sup>-</sup> est attribué aux souches produisant au moins 1000 fois moins de vitamine B12 que la souche parente SC510 Rif<sup>r</sup>, figure 46. Il est ainsi observé que 18 de 20 ces insertions chromosomiques, sur les 27, conduisent à un phénotype Cob<sup>-</sup> chez P. denitrificans comme cela est montré sur la figure 46.

Les insertions 19, 32, 24, 27, 37, 39, 26, 11 et 14 sont cartographiées dans la phase ouverte 22 (voir figure 46). Toutes ces insertions sont complémentées par le plasmide pXL618 qui ne contient que la phase 25 ouverte 22. Nous en déduisons que la phase ouverte 22 correspond à un gène cob que nous appelons cobQ. Aucune insertion n'a été obtenue dans la phase ouverte 23; cependant le plasmide pXL623 qui ne contient que cette phase ouverte (voir figure 46), complémente deux mutants cob d'Agrobacterium tumefaciens, G642 et G2043 (Cameron et al., 1989). La 30 phase ouverte 23 correspond donc à un gène cob nommé cobP. Les insertions 23, 13, 12, 30, 22, 40, 35, 10 et 17 qui sont cartographiées dans les phases ouvertes 24 et 25 entraînent un phénotype Cob<sup>-</sup> chez SC510 Rif<sup>r</sup>. Il semble donc qu'il s'agisse de deux phases ouvertes dont le produit est impliqué dans la biosynthèse des cobalamines. Toutefois 35 on ne peut exclure que ces insertions aient des effets polaires sur les

gènes positionnés en 3' tels que cobO. Il convient donc d'étudier la complémentation de ces mutants afin de voir si leur phénotype Cob- ne résulte pas d'un effet polaire.

Les mutants Cob d'Agrobacterium tumefaciens G622, G623 et G630 complémentés par le pXL156 ont été étudiés. Ces mutants ne sont pas complémentés par le plasmide pXL189 (Cameron et al., 1989) qui contient cobO comme seul gène cob. Au contraire ils sont complémentés par le plasmide pXL1908 qui contient cobO et la phase ouverte 25 en plus des phases ouvertes 27 à 30 (voir figure 45). Ces dernières ne peuvent être responsables de la complémentation de ces mutants puisque les protéines pour lesquelles elles codent n'interviennent pas dans la voie du coenzyme B12. Les complémentations observées ne peuvent donc être que le seul fait de la phase ouverte 25. De plus les mutants SC510 Rif<sup>R</sup> Tn5Sp cartographiés dans cette même phase ouverte (il s'agit des mutants 22, 40, 35, 10 et 17) sont complémentés par le plasmide pXL1908, voir figure 46, (portant cobO et la phase 25) alors qu'au moins deux d'entre eux ne sont pas complémentés par le pXL189 qui ne contient que cobO comme gène cob. Ces résultats montrent clairement que la phase ouverte 25 est un gène cob; ce gène cob est nommé cobN.

Les mutants SC510 Rif<sup>R</sup> Tn5Sp 23, 13 et 12, qui ont le phénotype Cob-, sont cartographiés dans la phase ouverte 24. Ces mutants ne sont pas complémentés par le plasmide pXL623 qui ne contient que le gène cobP. Au contraire ces mutants sont complémentés par le plasmide pXL593 qui contient cobP et la phase ouverte 24 ce qui indique que la phase ouverte 24 est responsable de leur complémentation. La phase ouverte 24 est donc un gène cob qui est nommé cobW.

#### EXEMPLE 5 - Gènes et protéines.

##### 5.1 - Fragment de 5,4 kb

Cinq gènes (cobA, cobB, cobC, cobD et cobE) sont donc définis sur le fragment ClaI-HindIII-HindIII-HindIII de 5,4 kb. Ils codent respectivement pour les protéines COB suivantes: COBA, COBB, COBC, COBD et COBE. Les parties codantes des gènes (cobA à cobE) sont

décrivées à la figure 15, ainsi que les séquences des protéines COBA à COBE. Des propriétés de chacune de ces protéines sont aussi présentées (composition en acides aminés, point isoélectrique, index de polarité et profil d'hydrophilicité).

5

#### 5.2 - Fragment de 8,7 kb

Huit gènes sont donc définis sur le fragment de 8,7 kb. Ces gènes cobF à cobM, codent respectivement pour les protéines COB suivantes : COBF, COBG, COBH, COBI, COBJ, COBK, COBL, COBM. Les parties codantes des gènes (cobF à cobM) sont décrites sur la figure 16, ainsi que les séquences des protéines COBF à COBM. Des propriétés de chacune de ces protéines sont aussi présentées (composition en acides aminés, poids moléculaire, point isoélectrique, index de polarité et profil d'hydrophilicité).

15

#### 5.3 - Fragment de 4,8 kb

Trois gènes (cobX, cobS, cobT) sont définis sur le fragment SalI-SalI-SalI-SalI-BglI de 4,8 kb. Ils codent respectivement pour les protéines suivantes: COBX, COBS et COBT. Les parties codantes de ces gènes sont décrites sur la figure 40, ainsi que les séquences des protéines COBX, COBS et COBT. Arbitrairement, l'ATG à la position 1512 de cobS a été choisi comme codon d'initiation plutôt que celui situé à la position 1485 (voir figure 32). Des propriétés de chacune de ces protéines sont aussi représentées (composition en acides aminés, point isoélectrique, index de polarité et profil d'hydrophobicité). COBT présente une poche hydrophile correspondant aux acides aminés 214 à 305.

20

25

#### 5.4 - Fragment de 3,9 kb

Deux gènes (cobU et cobV) sont définis sur le fragment SstI-SstI-BamHI de 3,9 kb. Ils codent respectivement pour les protéines suivantes: COBU et COBV. Les parties codantes de ces gènes

sont décrites sur la figure 41, ainsi que les séquences des protéines COBU à COBV. Des propriétés de chacune de ces protéines sont aussi représentées (composition en acides aminés, point isoélectrique, index de polarité et profil d'hydrophobicité).

5

### 5.5 - Fragment de 13,4 kb

Cinq gènes cob sont définis sur le fragment de 13.4 kb (cobQ, cobP, cobW, cobN et cobO et cobV). Ils codent respectivement pour les protéines suivantes: COBQ, COBP, COBW, COBN et COBO. Les parties codantes de ces gènes (cobQ, cobP, cobW, cobN et cobO) sont décrites 10 sur la figure 46, ainsi que les séquences des protéines COBQ, COBP, COBW, COBN et COBO. Des propriétés de chacune de ces protéines sont aussi représentées (composition en acides aminés, point isoélectrique, index de polarité et profil d'hydrophobicité).

D'après les profils d'hydrophilicité, qui ont été réalisés 15 suivant les programmes de Hopp et Woods (1981), toutes les protéines COB, à l'exception de COBV, sont vraisemblablement des protéines solubles, par opposition à des protéines membranaires, puisque l'on constate l'absence de grands domaines hydrophobes. COBV est soit une protéine membranaire, puisque l'on constate 4 longs domaines 20 hydrophobes (voir figure 41) soit une protéine cytoplasmique ayant d'importants domaines hydrophobes.

Pour toutes les séquences en acides aminés des protéines COB, il est indiqué comme premier acide aminé en position NH<sub>2</sub>-terminale une méthionine. Il est entendu que celle-ci peut être 25 excisée in vivo (Ben Bassat et Bauer, 1984). On sait que des règles concernant l'excision in vivo de la méthionine NH<sub>2</sub>-terminale par la méthionine aminopeptidase ont été proposées (Hirel et al., 1989).

Par ailleurs, ces séquences protéiques ont été comparées aux protéines de Genpro, qui est une extraction protéique de Genbank 30 (version 59) augmentée des parties codantes putatives supérieures à

200 acides aminés, suivant le programme de Kanehisa (1984). Aucune homologie significative n'a pu être mise en évidence avec les paramètres utilisés sur la version 59 de Genbank, sauf pour COBT. En effet, la protéine COBT présente un "core d'acide aminés acides" entre 5 les positions (en acides aminés) 224 et 293 (voir figure 40); dans cette portion de la protéine, plus d'un acide aminé sur 2 est un résidu acide glutamique ou aspartique; ce noyau d'acides aminés acides rend la protéine homologue sur cette région, suivant le programme de Kanehisa (1984), à d'autres protéines ayant elles aussi un tel noyau 10 acide. Les protéines les plus homologues sont : la protéine GARP de Plasmodium falciparum (Triglia et al., 1988), la troponine T cardiaque du rat (Jin et Lin, 1989), la prothymosine humaine et de rat (Eschenfeld et Berger, 1986), une protéine du rat androgène-dépendante 15 se liant à la spermine (Chang et al., 1987), les protéines "mid-size neurofilament subunit" humaine, de rat et de poulet (Myers et al., 1987, Levy et al., 1987, Zopf et al., 1987). La fonctionnalité de ces noyaux riches en résidus acides n'est pas connue; cependant ce noyau acide devrait soit permettre la fixation de cation métalliques tels que le Co<sup>++</sup>, ce qui rendrait à la protéine COBT un rôle de 20 métallothionéine à cobalt, ou bien permettre des interactions avec d'autres protéines.

#### EXEMPLE 6 - Etudes enzymatiques

##### 6.1 - Identification de protéines COB et de leurs gènes à partir d'activités enzymatiques

25 Cet exemple décrit comment, à partir d'une protéine purifiée, après avoir établi sa séquence NH<sub>2</sub>-terminale, il est possible de trouver le gène de structure correspondant parmi des gènes cob séquencés.

###### 6.1.1. Identification de la protéine COBA codée par le 30 gène cobA

La purification de la SUMT de Pseudomonas denitrificans a

été décrite (F. Blanche et al., 1989). La séquence NH<sub>2</sub>-terminale de la protéine ainsi purifiée a pu être réalisée selon la technique décrite ci-dessus. Les dix premiers acides aminés ont été identifiés :

	1	2	3	4	5	6	7	8	9	10
5	Met	Ile	Asp	Asp	Leu	Phe	Ala	Gly	Leu	Pro

La séquence NH<sub>2</sub>-terminale de la protéine COBA (figure 15) correspond exactement à cette séquence. Le poids moléculaire de la SUMT purifiée estimé par électrophorèse PAGE-SDS à 12,5 % est de 30 000. La protéine COBA a un poids moléculaire déduit de sa séquence de 10 29 234 (figure 15). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires indiquent clairement que la protéine COBA correspond à la SUMT. Le gène cobA est le gène de structure de la SUMT.

#### 6.1.2. Identification de la protéine COBB codée par le 15 gène CobB

##### a) Dosage de l'activité acide cobyrinique a,c-diamide synthase

Cet exemple illustre le dosage d'une activité de la voie de biosynthèse des corrinoïdes qui n'a encore jamais été décrite. Il s'agit de l'acide cobyrinique a,c-diamide synthase (ACDAS) qui 20 catalyse l'amidation de deux fonctions acides carboxyliques du noyau corrine ou descobaltocorrine aux positions a et c (figure 17). Le donneur de groupement NH<sub>2</sub> est la L-glutamine et la réaction consomme 1 molécule d'ATP par amidation de chaque fonction acide carboxylique. Le dosage qui est décrit ci-dessous s'applique à la réaction de 25 diamidation de l'acide cobyrinique; avec quelques modifications (détection en CLHP à 330 nm en particulier) il s'applique à la réaction de diaminidation de l'acide hydrogénobyrinique.

Le mélange d'incubation (250 µl de tris-HCl 0,1 M pH 7,6) contenant ATP (1 mM), MgCl<sub>2</sub> (2,5 mM), glutamine (1 mM), acide 30 cobyrinique (25 µM) ou acide hydrogénobyrinique (5 µM), cobyrinique a,c-diamide synthase (environ 1 unité d'activité) est incubé durant 1 heure à 30°C. A la fin de l'incubation, 125 µl d'une solution aqueuse

de KCN (2,6 g/l) et 125 µl d'HCl 0,2 M sont ajoutés au mélange qui est ensuite chauffé à 80°C pendant 10 minutes puis centrifugé 5 minutes à 5 000 g. 50 µl du surnageant de centrifugation sont analysés en CLHP. Ils sont injectés sur une colonne Nucleosil 5-C<sub>18</sub> de 25 cm et élués avec un gradient de 0 à 100 % de tampon B dans A en 30 minutes ; tampon A : phosphate de potassium 0,1 M pH 6,5, 10 mM KCN ; tampon B : phosphate de potassium 0,1 M pH 8, 10 mM KCN/acetonitrile (1/1). Les corrinoïdes sont détectés grâce à leur absorption UV à 371 nm. L'unité d'activité enzymatique est définie comme la quantité d'enzyme nécessaire pour synthétiser 1 nmole de groupements amides par heure dans les conditions décrites.

b) Purification de l'activité acide cobyrinique a,c-diamide synthase de Pseudomonas denitrificans

Cette expérience illustre comment une protéine de Pseudomonas denitrificans intervenant dans la voie de biosynthèse des cobalamines peut être purifiée.

A partir du dosage décrit à l'exemple 6.1.2 a), la purification de l'acide cobyrinique a,c-diamide synthase de Pseudomonas denitrificans est réalisée comme décrit ci-dessous.

Dans une expérience typique de purification, 7 g de cellules humides de la souche SC 510 Rif<sup>r</sup> dans laquelle on a introduit le plasmide pXL1500 (voir exemple 4.1. pour la description du pXL1500 ainsi que la figure 12) sont suspendues dans 30 ml de Tris-HCl 0,1 M pH 7,7 et soniquées durant 15 minutes à 4°C. L'extrait brut est ensuite récupéré par centrifugation 1 heure à 50 000 g puis 10 ml de cet extract sont injectés sur une colonne de Mono Q HR 10/10 équilibrée avec le même tampon. Les protéines sont éluées avec un gradient linéaire de KCl (0 à 0,5 M). Les fractions contenant l'activité enzymatique sont regroupées et concentrées à 2,5 ml. Après dilution avec 1 ml de Tris-HCl 25 mM pH 7,7, les protéines sont fractionnées sur une Mono Q HR 5/5 en utilisant le gradient de KCl (0 à 0,5 M) précédent. Les fractions actives sont regroupées, 1 ml de Tris-HCl 0,1 M pH 7,7 contenant 1,7 M de sulfate d'ammonium est ajouté à l'échantillon qui est ensuite chromatographié sur une colonne de

Phényl-Superose (Pharmacia) avec un gradient décroissant de sulfate d'ammonium (1,0 M à 0 M). Les fractions contenant l'activité recherchée sont rassemblées et chromatographiées sur une colonne Bio-Gel HPHT (Bio-Rad) avec un gradient de phosphate de potassium (0 à 5 0,35 M).

Après cette étape, l'enzyme est pure à plus de 95 %. Elle ne présente aucune protéine contaminante en PAGE-SDS. La pureté de la protéiné est confirmée par l'unicité de la séquence NH<sub>2</sub>-terminale. Son poids moléculaire dans cette technique est de 45 000. Les différentes 10 étapes de purification de l'ACDAS, avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous.

Tableau : Purification de l'ACDAS

	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification <sup>1</sup>
16	Extrait brut	10	200	8.5	-	-
	MonoQ 10/10	12	15.1	108	96	12.7
	MonoQ 5/5	3	3.75	272	60	32
20	Phényl-Superose	1	0.865	850	43	100
	Bio-Gel HPHT	2	0.451	1320	35	155

<sup>1</sup>ce facteur est calculé d'après l'augmentation de l'activité spécifique des fractions au cours de la purification.

25 c) Séquence NH<sub>2</sub>-terminale de l'acide cobyrinique a,c-diamide synthase de Pseudomonas denitrificans et identification du gène de structure de Pseudomonas denitrificans codant pour cette activité

Cet exemple illustre comment la séquence NH<sub>2</sub>-terminale d'une protéine qui intervient dans la voie de biosynthèse des cobalamines permet d'identifier le gène de structure qui code pour

cette protéine.

La séquence NH<sub>2</sub>-terminale de l'acide cobyrinique a,c-diamide synthase de Pseudomonas denitrificans purifiée comme décrit à l'exemple 6.1.2 b) a été réalisée comme décrit précédemment.

5 15 résidus ont été identifiés :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ser	Gly	Leu	Leu	Ile	Ala	Ala	Pro	Ala	Ser	Gly	Ser	Gly	Lys	Thr

10 La séquence NH<sub>2</sub>-terminale de la protéine COBB (figure 15) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 15, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminoterminale est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de l'ACDAS purifiée estimé par électrophorèse PAGE-SDS à 15 12,5 % est de 45 000. La protéine COBB a un poids moléculaire déduit de sa séquence de 45 676 (figure 15). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires indiquent clairement que la protéine COBB correspond à l'ACDAS. Le gène cobB est le gène de structure de l'ACDAS.

20 6.1.3. Identification de la protéine COBI codée par le gène CobI

a) Dosage d'une activité S-adénosyl-L-méthionine:précorrine-2 méthyl-transférase

25 Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des corronoïdes qui n'a encore jamais été décrite. Il s'agit de la S-adénosyl-L-méthionine:-2 méthyltransférase (SP<sub>2</sub>MT) qui catalyse le transfert d'un groupement méthyl de la S-adénosyl-L-méthionine (SAM) au précorrine-2 pour donner le précorrine-3 (figure 18). Les facteurs II et III, produits d'oxydation 30 respectivement du précorrine-2 et du précorrine-3 ont déjà été purifiés à partir d'extraits cellulaires de Propionibacterium shermanii (Battersby et MacDonald, 1982, Scott et al., 1984); le

précorrine-2 et le précorrine-3 sont reconnus comme des intermédiaires présumés de biosynthèse du coenzyme B<sub>12</sub> mais ils n'ont jamais été purifiés. Pour cette raison, l'activité correspondante n'a jamais été ni dosée, ni purifiée auparavant. Le substrat de la réaction enzymatique, précorrine-2, est une molécule très labile qu'il n'est pas possible de conserver car elle s'oxyde spontanément en présence de traces même infimes d'oxygène (Battersby et MacDonald, 1982). Le principe de ce test enzymatique repose donc sur la possibilité de générer extemporanément, à l'aide d'un extrait enzymatique de la souche SC510 Rif<sup>r</sup> dans laquelle on a introduit le plasmide pXL1500, le précorrine-2 à partir du SAM et d'acide δ-aminolévulinique. L'incubation doit être effectuée en conditions de stricte anaérobie.

Les fractions contenant la SP<sub>2</sub>MT sont incubées dans 1 ml de Tris-HCl 0,1 M pH 7,7 en présence de 5 mM DTT, 1 mM EDTA, 100 μM [methyl-<sup>3</sup>H]-SAM (1 μCi), 0,8 mM acide δ-aminolévulinique et 6 mg d'extrait enzymatique brut de la souche de Pseudomonas denitrificans SC510 Rif<sup>r</sup> pXL1500 pendant 3 heures à 30°C. La souche SC510 Rif<sup>r</sup> pXL1500 contient une forte activité SUMT (F. Blanche et al., 1989). Les composés tétrapyrroliques produits durant l'incubation sont fixés sur une colonne d'échangeur d'anions DEAE-Sephadex et estérifiés dans le méthanol à 5 % d'acide sulfurique en l'absence d'oxygène. Les dérivés diméthylés et triméthylés de l'uro'gen III sont ensuite séparés par chromatographie sur couche mince de silice en utilisant le dichlorométhane/méthanol (98,3/1,7) comme système éluant (F. Blanche et al., 1989). L'activité de la SP<sub>2</sub>MT est exprimée par le rapport de la quantité de dérivés triméthylés obtenus sur l'ensemble des dérivés (di- et tri-) méthylés produits ramenée, à la quantité de protéine. L'extrait de SC510 Rif<sup>r</sup> pXL1500 introduit dans le test ne présente pas d'activité SP<sub>2</sub>MT détectable dans les conditions de dosage (le ratio précorrine-3 produits sur précorrine-2 produits durant le test est inférieur à 0,05).

b) Purification de la S-adénosyl-L-méthionine:précorrine-2 méthyl-transférase de Pseudomonas denitrificans

Cette expérience illustre comment une protéine de

Pseudomonas denitrificans intervenant dans la voie de biosynthèse de cobalamines peut être purifiée lorsqu'un dosage de l'activité en question existe.

La protéine est purifiée à partir de cellules de SC510 5  $\text{Rif}^r$  contenant le plasmide pXL253. Il s'agit du plasmide pKT230 sur lequel a été inséré le fragment EcoRI de 8,7 kb (figure 13). Dans une expérience typique de purification, 50 g de cellules humides de la souche SC150  $\text{Rif}^r$  dans laquelle a été introduit le plasmide pXL253 sont suspendues dans 250 ml de phosphate de potassium 0,1M pH 7,7, 5 mM DTT 10 et soniquées durant 15 minutes à 4°C. Après centrifugation à 50 000 g durant 1 heure, le surnageant est passé à travers une colonne de DEAE-Sephadex (10 ml de gel) pour éliminer les composés tétrapyrroliques. Le pH de l'extrait brut ainsi obtenu est ajusté à pH 15 7,7 avec KOH 0,1 M. Les protéines précipitant entre 33 % et 45 % de saturation en sulfate d'ammonium sont collectées et dissoutes dans 40 ml de Tris-HCl 0,1M pH 7,7, 5 mM DTT. Cette solution est passée à travers une colonne de Sephadex G-25 éluée avec du Tris-HCl 10 mM pH 7,7, 5 mM DTT et les protéines collectées sont injectées sur une colonne de DEAE-Trisacryl-M. Les protéines sont éluées avec un 20 gradient linéaire de 0 à 0,25 M KCl et les fractions contenant l'activité SP<sub>2</sub>MT sont regroupées et passées une seconde fois à travers une colonne de Sephadex G-25 comme ci-dessus. La fraction protéique est injectée sur une colonne d'Ultrogel HA (IBF) équilibrée dans le Tris-HCl 10 mM pH 7,7, 5 mM DTT. Les protéines sont éluées avec un 25 gradient linéaire de 0 à 50 mM de phosphate de potassium pH 7,8 contenant 5 mM DTT. Les fractions contenant l'activité recherchée sont regroupées et injectées sur une MonoQ HR 5/5 (Pharmacia) équilibrée avec du Tris-HCl 50 mM pH 7,7, 5 mM DTT. La SP<sub>2</sub>MT est éluée avec un gradient linéaire (0 à 0,25 M) de KCl. A la sortie de l'étape MonoQ, 30 l'électrophorèse en PAGE-SDS (12,5 %) avec coloration aux sels d'argent révèle que l'enzyme est pure à plus de 99 %. Ceci est confirmé par l'unicité de la séquence NH<sub>2</sub>-terminale de la protéine. Le poids moléculaire calculé à partir de l'électrophorèse en conditions dénaturantes (PAGE-SDS à 12,5 %) est de 26 500. Les étapes de 35 purification de la SP<sub>2</sub>MT avec leurs rendements, sont décrites dans le

tableau ci-dessous.

Tableau : Purification de la SP<sub>2</sub>MT

	Etape de purification	Vol (ml)	Protéines (mg)	Facteur de purification <sup>1</sup>
5	Extrait brut	300	6000	-
	Précipitation (33-45 %)	40	1530	3.9
	DEAE-Trisacryl-M	57	355	16.9
	Ultrogel HA	30	71	85
10	MonoQ HR 5/5	12	33.5	179

<sup>1</sup>ce facteur est calculé d'après le rendement en protéines.

c) Séquence NH<sub>2</sub>-terminale de la SP<sub>2</sub>MT et identification du gène de structure codant pour cette activité

Cet exemple illustre comment la séquence NH<sub>2</sub>-terminale d'une protéine qui intervient dans la voie de biosynthèse permet d'identifier le gène de structure qui code pour cette protéine. Dans l'exemple présent, il s'agit du gène de structure de la SP<sub>2</sub>MT.

La séquence NH<sub>2</sub>-terminale de la protéine purifiée a été réalisée comme décrit précédemment. Les 15 premiers acides aminés ont été identifiés :

1    2    3    4    5    6    7    8    9    10    11    12    13    14    15

Ser Gly Val Gly Val Gly Arg Leu Ile Gly Val Gly Thr Gly Pro

La séquence NH<sub>2</sub>-terminale de la protéine COBI (figure 16) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 16 une méthionine précède la séquence peptidique déduite de la séquence nucléotidique. Il en résulte que la méthionine aminoterminale est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids

moléculaire de la SP<sub>2</sub>MT purifiée estimé par électrophorèse PAGE-SDS à 12,5 % est de 26 500. La protéine COBI a un poids moléculaire déduit de sa séquence en acides aminés de 25 878 (figure 16). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires indiquent clairement que la protéine COBI correspond à la SP<sub>2</sub>MT. Le gène cobi est le gène de structure de la SP<sub>2</sub>MT.

6.1.4. Identification de la protéine COBH codée par le gène cobH

a) Dosage de l'activité precorrin-8x mutase.

Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été décrite. Il s'agit de la précorrin-8x mutase. Cette enzyme catalyse le transfert du groupement méthyl de la position C-11 à la position C-12 lors de la transformation du précorrin-8x en acide hydrogénobyrinique (voir la nomenclature des carbones FIG. 19; PL. 68). Plus généralement, c'est l'enzyme catalysant le transfert du groupement méthyl de C-11 vers C-12, conduisant ainsi au noyau corrine. L'enzyme est ici appelée mutase, bien qu'il ne soit pas formellement démontré que le transfert du groupement méthyl soit intramoléculaire, même si cela est très vraisemblable.

L'activité enzymatique est mise en évidence par la transformation du précorrin-8x (5 µM) en acide hydrogénobyrinique au cours d'incubations en présence de fractions enzymatiques dans du Tris-HCl 0.1 M pH 7.7, 1 mM EDTA, à 30 °C durant 1 h. En fin d'incubation, la réaction est stoppée par chauffage à 80 °C durant 10 min et après centrifugation à 3000 × g durant 10 min, l'acide hydrogénobyrinique formé présent le surnageant est analysé par CLHP (Cf exemple 6.1.2.a).

b) Purification de la précorrin-8x mutase.

La purification de la précorrin-8x mutase de *Pseudomonas denitrificans*

est réalisée comme décrit ci-dessous.

Durant cette purification, toutes les solutions tampons sont ajustées à pH 7.7.

Dans une expérience typique de purification, 50 g de cellules de la souche SC510 Rifr portant le plasmide pXL253 (plasmide pKT230 sur lequel a été cloné, au site EcoRI, le fragment de 8.7 kb, figure 13), obtenues après culture en milieu PS4, sont resuspendues dans 200 ml de tampon phosphate de potassium 0.1 M, et soniquées durant 12 min. Après centrifugation à 50 000 x g durant 1 heure, le surnageant est passé à travers une colonne de DEAE-Sephadex (10 ml de gel) pour éliminer les composés tetrapyrroliques. Le pH de la solution est immédiatement ajusté à 7.7 avec une solution 1 M KOH. La fraction protéique précipitant entre 40 et 60% de saturation en sulfate d'ammonium est collectée par centrifugation et dissoute dans 50 ml de Tris-HCl 0.1 M. Cet échantillon est ensuite injecté sur une colonne d'Ultrogel AcA 54 (IBF, France)(volume de gel 1000 ml), et les protéines sont éluées à un débit de 60 ml/h avec du Tris-HCl 50 mM. Les fractions contenant l'activité sont regroupées et injectées sur une colonne de DEAE-Trisacryl M (IBF, France) équilibrée avec du Tris-HCl 50 mM, et les protéines sont éluées avec un gradient de 0 à 0.2 M KCl. Les fractions contenant la protéine à purifier sont regroupées, passées à travers une colonne de Sephadex G25 équilibrée en Tris-HCl 10 mM. La fraction protéique est injectée sur une colonne d'Ultrogel HA (IBF, France) équilibrée avec du Tris-HCl 10 mM, et les protéines sont éluées avec un gradient de 0 à 0.1 M de phosphate de potassium, puis la fraction active est chromatographiée sur une colonne de Phenyl-Sepharose CL 4B (Pharmacia) dans du phosphate de potassium 10 mM, éluée avec un gradient de 0.65 à 0 M de sulfate d'ammonium. Les fractions actives sont regroupées. La protéine ainsi obtenue est pure à plus de 95% (d'après les résultats d'électrophorèse PAGE-SDS à 12.5% et coloration aux sels d'argent). La pureté de la protéine est confirmée par l'unicité de la séquence N-terminale. Son poids moléculaire calculé à l'aide de cette technique est de 22 000. Les étapes de purification de la precorrin-8x mutase avec leurs

rendements de purification sont décrites sur le tableau ci-dessous.

tableau: Purification de la précorrin-8x mutase.

	Etape de purification	Vol (ml)	Protéines (mg)	Facteur de purification <sup>1</sup>
5	Extrait brut	250	6000	-
	Précipitation (40-60 %)	50	2350	2.6
	Ultrogel ACA 54	70	655	9.2
	DEAE-Trisacryl M	30	271	22
	Ultrogel HA	22	93	65
10	Phényl Sépharose	12	31	194

<sup>1</sup>ce facteur est calculé d'après le rendement en protéines

c) Séquence NH<sub>2</sub>-terminale de la précorrin-8x mutase et identification de son gène de structure.

15 Cet exemple illustre comment la séquence NH<sub>2</sub>-terminale d'une protéine qui intervient dans la voie de biosynthèse permet d'identifier le gène de structure qui code pour cette protéine.

La séquence NH<sub>2</sub>-terminale de cette protéine a été réalisée comme décrit précédemment. 15 résidus ont été identifiés :

20        1    2    3    4    5    6    7    8    9    10   11   12   13   14   15  
          Pro Glu Tyr Asp Tyr Ile Arg Asp Gly Asn Ala Ile Tyr Glu Arg

25 La séquence NH<sub>2</sub>-terminale de la protéine COBH (figure 16) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 16 une méthionine précède la séquence peptidique déterminée par le séquençage décrit ci-dessus. Il en résulte que la méthionine aminoterminal est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le

deuxième résidu étant une proline, cette excision est conforme aux règles déjà énoncées (Hirel et al., 1989). Le poids moléculaire de la precorrin-8x mutase purifiée estimé par électrophorèse PAGE-SDS à 12,5 % est de 22 000. La protéine COBH a un poids moléculaire déduit de sa 5 séquence de 22 050 (figure 16). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires, de ces protéines, indiquent clairement que la protéine COBH correspond à la precorrin-8x mutase. cobH est le gène de structure de la precorrin-8x mutase.

d) Préparation, isolement et identification du precorrin-8x.

10 Dans une expérience typique de préparation du precorrin-8x, un extrait enzymatique brut de la souche SC510 Rifr pXL253 (1000 mg de protéines) est incubé en anaérobiose durant 20 h à 30°C dans 100 ml de tampon Tris-HCl 0,1 M pH 7,7 avec 1000 nmol de trimethylisobacteriochlorin préparé comme décrit précédemment 15 (Battersby et al., 1982), EDTA (1 mM), ATP (100 µmol), MgCl<sub>2</sub> (250 µmol), NADH (50 µmol), NADPH (50 µmol), SAM (50 µmol) et acide hydrogénobyrinique (20 µmol). En fin d'incubation, le precorrin-8x est le produit tétrapyrrrolique formé majoritaire. Il est isolé et purifié par CLHP sur une colonne µBondapak C18 (Waters) en utilisant 20 un gradient d'élution linéaire de 0 à 50% d'acétonitrile dans un tampon phosphate de potassium, pH 5,8. La masse du precorrin-8x (*m/z* = 880) et la masse de son dérivé methyl ester (*m/z* = 978) indiquent qu'il s'agit d'un composé ayant la même formule brute que l'acide hydrogénobyrinique. Les caractéristiques UV/Visible et de 25 fluorescence sont très différentes de celles de l'acide hydrogénobyrinique et indiquent que la molécule possède deux chromophores séparés. La seule réaction enzymatique d'isomérisation entre precorrin-6x (Thibaut et al., 1990) et l'acide hydrogénobyrinique étant la migration du méthyl de C-11 vers 30 C-12, le precorrin-8x est le dernier intermédiaire avant l'acide hydrogénobyrinique et la réaction correspondante est la migration du méthyl de C-11 vers C-12, catalysée par la precorrin-8x mutase.

#### 6.1.5. Identification de la protéine COBU codée par le

**gène cobU****a) Dosage de l'activité**

nicotinate-nucleotide:dimethylbenzimidazole phosphoribosyltransferase (figure 5, réaction 5). Cet exemple illustre le dosage d'une activité enzymatique directement liée à la voie de biosynthèse des cobalamines. Il s'agit de la nicotinate-nucleotide:dimethylbenzimidazole phosphoribosyltransferase (NN:DMBI PRT) (EC 2.4.2.21). Les fractions contenant l'activité NN:DMBI PRT (environ 5 unités) sont incubées à 30°C durant 8 mn dans 500 µl de tampon glycine:NaOH 0.1 M pH 9.7 en présence de 1 mM NaMN (acide nicotinique mononucéotide) et 10 µM DMBI. La réaction est ensuite stoppée par chauffage à 80°C durant 10 mn, le mélange réactionnel est dilué par 4 volumes d'eau et 100 µl de cette solution sont injectés sur une colonne CLHP Nucleosil 5-C8 de 15 cm éluée avec un mélange 0.1 M phosphate de potassium pH 2.9:acétonitrile (93:7) à un débit de 1 ml/min. Le α-ribazole 5'-phosphate est détecté et quantifié par fluorimétrie (excitation : 260 nm; émission >370 nm). L'unité d'activité enzymatique est définie comme la quantité d'enzyme nécessaire pour générer 1 nmol de α-ribazole 5'-phosphate par heure dans ces conditions.

b) Purification de l'activité NN:DMBI PRT de *Pseudomonas denitrificans* Cette expérience illustre comment une protéine de *P. denitrificans* intervenant dans la voie de biosynthèse des cobalamines peut être purifiée. A partir du dosage décrit dans l'exemple 6.1.5.a), la purification de la NM:DMBI PRT de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous. Dans une expérience typique de purification on utilise 10 g de cellules humides de la souche SC510 Rifr dans laquelle le plasmide pXL1490B a été introduit comme cela est décrit précédemment. Le plasmide pXL1490B est décrit sur la figure 38; ce plasmide a été obtenu par clonage du fragment BamHI-SstI-SstI de 3.85 kb du pXL519 (voir figure 38). Ce plasmide porte donc les gènes cobU et cobV de *P. denitrificans*. Les cellules cultivées en milieu PS4, supplémenté en lividomycine, comme cela est décrit précédemment, sont récoltées après 96 heures de culture en milieu PS4. Elles sont resuspendues dans 25 ml de tampon 0.1M Tris-HCl pH 7.2 et soniquées

durant 15 mn à 4°C. L'extrait brut est ensuite récupéré par centrifugation durant 1 h à 50 000 g puis passé à travers une colonne de DEAE-Trisacryl M (IBF, France) équilibrée avec le même tampon. 10 % de l'éluat (120 mg de protéines) est fractionné sur une colonne de 5 Mono Q HR 10/10 en utilisant un gradient de KCl (de 0 à 0.6 M). Les fractions actives sont regroupées et concentrées à 2 ml par ultrafiltration puis après mélange avec un volume de tampon Tris-HCl 30 mM pH 7.2, l'échantillon est fractionné une seconde fois sur une Mono Q HR 5/5 comme précédemment. Les fractions actives sont 10 regroupées puis l'échantillon est amené à une molarité de 1 M à l'aide de sulfate d'ammonium et chromatographié sur une colonne de Phényl-Superose HR 5/5 élué avec un gradient décroissant de sulfate d'ammonium (de 1 M à 0 M). Les fractions contenant l'activité recherchée sont rassemblées, concentrées par ultrafiltration, et 15 chromatographiées sur une colonne de perméation de gel Bio-Sil 250, éluée avec du phosphate de sodium 20 mM-sulfate de sodium 50 mM pH 6.8.

Après cette étape, l'enzyme est pure à plus de 95 %. Elle ne présente 20 aucune protéine contaminante en PAGE-SDS. Cette pureté est confirmée par l'unicité de la séquence NH<sub>2</sub>-terminale. Son poids moléculaire dans cette technique est de 35 000. Les différentes étapes de purification de la NN:DMBI PRT sont portées sur le tableau ci-dessous.

Tableau : Purification de la NN:DMBI PRT de *P. denitrificans*

	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification <sup>1</sup>
5	Extrait brut	6,0	120	2650	-	
	MonoQ 10/10	6,0	12,07	13515	51,3	5,1
	MonoQ 5/5	3,0	6,19	20140	39,2	7,6
	Phényl-Superose	1,5	2,60	35510	29,0	13,4
	Bio-Sil 250	1,2	1,92	39750	24,0	15,0

c) Séquence NH<sub>2</sub>-terminale de la NN:DMBI PRT de *P. denitrificans* et identification du gène de structure de *Pseudomonas denitrificans* codant pour cette activité. La séquence NH<sub>2</sub>-terminale de la NN:DMBI PRT de *Pseudomonas denitrificans* purifiée comme décrit dans l'exemple 6.1.5b) a été réalisée selon la technique décrite ci-dessus. Les 15 premiers résidus ont été identifiés :

15    1    2    3    4    5    6    7    8    9    10    11    12    13    14    15  
 Ser Ala Ser Gly Leu Pro Phe Asp Asp Phe Arg Glu Leu LeuArg

La séquence NH<sub>2</sub>-terminale de la protéine COBU (figure 41) correspond à cette séquence, si ce n'est que dans la séquence présentée sur la figure 41, une méthionine précède le premier acide aminé de la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminoterminale est certainement excisée *in vivo* par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de la N-transglycosidase purifiée estimé par électrophorèse PAGE-SDS à 12,5 % est de 35 000. La protéine COBU a un poids moléculaire déduit de sa séquence de 34 642 (figure 41). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires indiquent clairement que la protéine COBU correspond à la NN:DMBI PRT. Le gène cobU est le gène de structure de la NN:DMBI PRT.

d) Spécificité de la NN:DBI PRT pour de DBI. Cet exemple illustre comment l'étude de la spécificité de la NN:DMBI PRT de *P. denitrificans* permet de faire biosynthétiser à *P. denitrificans* diverses cobamides en utilisant les propriétés catalytiques de la 5 NN:DMBI PRT de *P. denitrificans* pour effectuer la synthèse de la base nucléotidique en question.

Le substrat de l'enzyme pour synthétiser des cobalamines est le 5,6-diméthylbenzimidazole. Le benzimidazole et le 5-méthylbenzimidazole, respectivement, sont des substrats de la 10 réaction avec des vitesses de réaction de 157 % et de 92 %, respectivement, comparé au substrat naturel (5,6-diméthylbenzimidazole), la concentration en NaMN étant fixée à 2 mM. La spécificité de la NN:DMBI PRT de *P. denitrificans* est donc faible pour les substrats à noyau benzimidazole. On peut donc utiliser 15 la souche de *P. denitrificans* SC510 Rifr (Cameron et al., 1989), la cultiver en milieu PS4 où le 5,6-diméthylbenzimidazole est remplacé par du benzimidazole ou du 5-diméthylbenzimidazole, respectivement, afin de faire synthétiser à la bactérie de la Co $\alpha$ -(benzimidazolyl)-Co $\beta$ -cyanocobamide, 20 Co $\alpha$ -(5-méthylbenzimidazolyl)-Co $\beta$ -cyanocobamide respectivement. Il est certain que d'autres cobamides pourraient être synthétisées de cette manière.

#### 6.1.6. Identification de la protéine COBV codée par le gène CobV.

25 Cet exemple illustre comment le dosage d'une activité de la voie de biosynthèse du coenzyme B12 chez *P. denitrificans*, puis la purification partielle de cette activité peut permettre d'identifier le gène de structure de cette enzyme chez *P. denitrificans*.

a) dosage de l'activité GDP-cobinamide: $\alpha$ -ribazole (5'-phosphate) 30 cobinamidephosphotransférase (ou cobalamine (5'-phosphate) synthase) Cet exemple illustre le dosage d'une activité directement liée à la voie de biosynthèse des cobalamines. Il s'agit de la cobalamine (5'-phosphate) synthase. Les fractions contenant l'activité (environ 5

à 10 unités) sont incubées à l'obscurité à 30°C dans 500 µl de tampon Tris-HCl 0.3 M pH 9.0 en présence de 1 mM EDTA, 12.5 mM MgCl<sub>2</sub>, 50 µM α-ribazole 5'-phosphate et 20 µM GDP-cobinamide [sous forme 5'-deoxy-5'-adénosyl (Ado) ou coenzyme]. Après 15 mn d'incubation, 500 5 µl de cyanure de potassium 20 mM sont ajoutés et la solution est chauffée à 80°C durant 10 mn. Après centrifugation pour éliminer le matériel précipité, la vitamine B<sub>12</sub> 5'-phosphate présente dans le surnageant est dosée comme cela est décrit à l'exemple 9. Une unité de cobalamine (5'-phosphate) synthase est définie comme la quantité 10 d'enzyme nécessaire pour générer 1 nmol de cobalamine 5'-phosphate par h dans les conditions décrites ci-dessus.

L'Ado-GDP-cobinamide est obtenu par incubation de l'Ado-cobinamide phosphate (Blanche et al., 1989) avec un extrait de SC510 Rifr pXL623 dans les conditions du dosage de la cobinamide 15 phosphate guanylyltransferase (voir 6.1.11.b). Le α-ribazole et le α-ribazole-5'-phosphate sont isolés à partir de cultures de SC510 Rifr et purifiés par CLHP dans les conditions du dosage décrit dans l'exemple 6.1.5a).

b) Purification partielle de la cobalamine (5'-phosphate) synthase.  
20 Cette expérience illustre comment une activité enzymatique de P. denitrificans intervenant dans la voie de biosynthèse des cobalamines de P. denitrificans peut être partiellement purifiée. A partir du dosage décrit ci-dessus, la purification de la cobalamine (5'-phosphate) synthase est réalisée. Pour ce faire, dans une 25 expérience typique de purification, 10 g de cellules humides de la souche SC510 Rifr dans laquelle le plasmide pXL1490B a été introduit comme cela est décrit précédemment. Le plasmide pXL1490B est décrit sur la figure 38; ce plasmide correspond au fragment SstI-SstI-BamHI de 3.85 kb cloné dans le pKT230. Ce plasmide porte les gènes cobU et 30 cobV de P. denitrificans. La présence de ce plasmide chez P. denitrificans SC510 Rifr conduit à une amplification de l'activité cobalamine (5'-phosphate) synthase d'un facteur 100 environ; il est donc probable que l'insert porté par le plasmide pXL1490B contient le 35 gène de structure de cette enzyme; ce gène ne peut donc être que cobU ou cobV. Les cellules SC510 Rifr pXL1490B sont obtenues par culture en

milieu PS4 supplémenté avec de la lividomycine comme cela est décrit ci-dessus. Les cellules sont centrifugées puis resuspendues dans 25 ml de tampon Tris-HCl 0.1 M (pH 8.3)-1 mM EDTA (tampon A) et soniqués pendant 15 mn à 4°C. L'extrait brut est ensuite récupéré par 5 centrifugation 1 h à 50 000 g et passé à travers une colonne de Sephadex G-25 équilibrée avec du tampon A. La fraction protéique est récupérée et injectée par fraction de 300 µl (7.5 mg de protéines) sur une colonne de Superose 12 HR 10/30 éluée dans du tampon A. La fraction exclue est récupérée, mélangée à un volume égal de tampon 10 A-1.0 M sulfate d'ammonium et chromatographiée sur une colonne de Phényl-Supérose HR 5/5. Les protéines sont éluées avec un gradient décroissant de sulfate d'ammonium (de 0.5 M à 0 M) dans du tampon A, suivi d'un plateau à 0 M de sulfate d'ammonium dans le but d'éluer l'activité cobalamine (5'-phosphate) synthase. La purification 15 partielle de cette enzyme est décrite dans le tableau ci-dessous, sur la base de 75 mg de protéines introduits au départ dans le processus de purification.

Tableau : Purification partielle de la cobalamine (5'-phosphate) synthase de *P. denitrificans*.

	Etape de purification	vol (ml)	Protéines (mg)	ACTIVITÉ SPÉ. (u/mg de protéines)	Rendement	Facteur de purification <sup>1</sup>
	Extrait brut	3.0	75	325	-	-
25	Superose 12HR	50.0	2.9	6 810	81	21
	Phenyl-Superose	4.5	0.35	17 850	26	55

c) Spécificité de la cobalamine (5'-phosphate) synthase. Le Km pour le (Ado)GDP-cobinamide est de 0.9 µM. Toutefois, l'enzyme présente la même affinité et une vitesse de réaction pratiquement identiques pour la forme (CN, aq) du substrat. Le Km de l'enzyme pour le α-ribazole

5'-phosphate est de 2.7  $\mu\text{M}$  environ. De plus, les préparations les plus pures de cobalamine (5'-phosphate) synthase catalysent la réaction du Ado-GDP-cobinamide avec le  $\alpha$ -ribazole pour donner le coenzyme B12 et dans ces conditions aucune accumulation de cobalamine 5'-phosphate n'est observée. Le  $K_m$  de l'enzyme pour le  $\alpha$ -ribazole est de 7.8  $\mu\text{M}$ . Des concentrations intracellulaires en  $\alpha$ -ribazole 5'-phosphate et en  $\alpha$ -ribazole de 30 et de 700  $\mu\text{M}$  respectivement, ont été mesurées par CLHP au cours de la production de cobalamines de *SC510 Rifr* en milieu PS4 dans les conditions de cultures décrites dans l'exemple 6.1.5a).

10 Ceci montre que le coenzyme B12 peut être généré directement à partir du Ado-GDP-cobinamide par la cobalamine (5'-phosphate) synthase sans l'intervention d'une cobalamine 5'-phosphatase.

L'absence d'accumulation ou la présence de traces de cobalamine 5'-phosphate dans les cultures de *P. denitrificans* SC510 Rifr confirme que le coenzyme B12 est produit par la réaction directe du Ado-GDP-cobinamide avec le  $\alpha$ -ribazole *in vivo*.

15 Cette réaction directe a déjà été observée et décrite *in vitro* chez *Propionibacterium shermanii* (Ronzio et al., 1967; Renz, 1968). Comme le gène de structure de la cobalamine (5'-phosphate) synthase ne peut être que *cobU* ou *cobV* puisque l'amplification chez *P. denitrificans* d'un fragment portant ces deux gènes *cob* de *P. denitrificans* conduit à une augmentation par un facteur 100 de l'activité cobalamine (5'-phosphate) synthase et que le gène *cobU* est le gène de structure de la NN:DMBI PRT, *cobV* est donc le gène de structure de la cobalamine (5'-phosphate) synthase.

#### 6.1.7. Identification de la protéine COBK codée par le gène *cobK*.

##### a) dosage de l'activité precorrin-6x réductase.

Cet exemple illustre le dosage d'une activité enzymatique nouvelle directement liée à la voie de biosynthèse des cobalamines. Il s'agit de la precorrin-6x réductase.

Les fractions contenant l'activité precorrin-6x réductase (environ 0.05 unités, U) sont incubées à 30°C durant 60 min dans 250 µl de tampon Tris-HCl 0.1 M pH 7.7 en présence de 1 mM EDTA, 500 µM NADPH, 25 µM [methyl-3H]SAM (80 µCi/µmol), 4 µM precorrin-6x (Thibaut 5 et al., 1990), et 0.5 U de dihydroprecorrin-6x méthylase partiellement purifiée (voir préparation ci-dessous). La réaction est ensuite stoppée par chauffage à 80°C durant 5 min et, après centrifugation à 5000 x g durant 5 min, le surnageant est injecté sur une colonne de DEAE-Sephadex (contenant 200 µl de gel). La colonne 10 est ensuite lavée extensivement avec le tampon Tris-HCl, et les composés fixés sont élués avec 4 ml de 1 M HCl. La radioactivité dans cet éluat est comptée en scintillation liquide. L'unité d'activité enzymatique est définie comme la quantité d'enzyme nécessaire pour réduire 1 nmol de precorrin-6x par h dans ces conditions.

15 La dihydroprecorrin-6x méthylase est partiellement purifiée à partir d'un extrait brut de SC510 Rifr pXL253, sur une colonne d'échange d'anions Mono Q HR 5/5 (Pharmacia). La colonne est éluée avec un gradient linéaire de 0 à 0.4 M de KCl dans du tampon 0.1 M Tris-HCl pH 7.7. L'activité enzymatique est éluée à 0.35 M KCl. Cette 20 activité est détectée et quantifiée grâce au test d'activité de la precorrin-6x reductase défini ci-dessus (en présence de 0.5 U de precorrin-6x reductase dans le milieu d'incubation). Après l'étape de Mono Q, les fractions contenant de l'activité dihydroprecorrin-6x méthylase sont totalement dépourvues 25 d'activité precorrin-6x réductase. L'unité d'activité méthylase est définie comme la quantité d'enzyme nécessaire pour transférer 1 nmol de groupements méthyls sur le dihydroprecorrin-6x par h dans les conditions décrites ci-dessus.

b) purification de l'activité precorrin-6x réductase.

30 A partir du dosage décrit ci-dessus, la purification de la precorrin-6x réductase de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous.

Dans une expérience typique de purification, 100 g de cellules humides

de la souche SC510 Rifr dans laquelle on a introduit le plasmide pXL253 (plasmide pKT230 sur lequel a été cloné, au site EcoRI, le fragment de 8.7 kb, figure 13) sont suspendues dans 200 ml de tampon Tris-HCl 0.1 M pH 7.7-1 mM EDTA (tampon A), et soniquées 5 durant 15 min à 4°C. L'extrait brut est ensuite récupéré par centrifugation durant 1 h à 50 000 x g et passé en trois fois à travers une colonne de Sephadex G 25 équilibrée avec le tampon A. Les trois fractions exclues du gel sont poolées et ajustées à 1 l avec le tampon A. Les protéines précipitant entre 25 et 40% de saturation 10 en sulfate d'ammonium sont collectées par centrifugation et resuspendues dans 50 ml de tampon A et cette solution est désalée à travers une colonne de Sephadex G 25 équilibrée avec du tampon B (25 mM Tris-HCl-500 µM DTT-15% glycérol). La solution de protéines est ensuite injectée à 2.5 ml/min sur une colonne de Q 15 Sepharose Fast Flow (Pharmacia) équilibrée avec du tampon B, et les protéines sont éluées avec un mélange tampon B-0.2 M KCL. Cette fraction est désalée sur une colonne de Sephadex G 25 équilibré avec du tampon C (50 mM Tris-HCl-500 µM DTT-15% glycérol). La solution protéique est ensuite fractionnée (100 mg de protéines à chaque 20 chromatographie) sur une colonne Mono Q HR 10/10 (Pharmacia) à l'aide d'un gradient de 0 à 0.4 M KCL dans le tampon C, puis la fraction contenant l'activité est chromatographiée sur une colonne Phenyl-Superose HR 10/10 (Pharmacia) dans un gradient linéaire décroissant de sulfate d'ammonium (de 1 à 0 M). La fraction active est 25 désalée et la precorin-6x réductase est repurifiée sur une colonne Mono Q HR 5/5. Elle est éluée dans le tampon 50 mM Tris-HCl pH 8.1-500 µM DTT-15% glycérol, avec un gradient de 0 à 0.2 M de KCL. Pour parfaire la purification, la protéine est finalement chromatographiée sur une colonne Bio-Sil 250 (Bio-Rad) éluée avec 30 20 mM potassium phosphate-50 mM sodium sulfate, pH 6.8-500 µM DTT-15 % glycérol. Après cette étape, l'enzyme est pure à plus de 95%. Elle ne présente aucune protéine contaminante en PAGE-SDS, les protéines étant révélées au nitrate d'argent. Ce degré de pureté est confirmé par l'unicité de la séquence NH<sub>2</sub>-terminale. Son poids 35 moléculaire dans cette technique est de 31 000. Les différentes étapes

de purification de la precorrin-6x réductase, avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous.

Tableau: Purification de la precorrin-6x reductase.

5	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification <sup>1</sup>
10	Extrait brut	270	9600	0.535	-	-
S.A. 25 40%	100	4160	1.14	92	2.1	
Q Sepharose	150	1044	3.64	74	6.8	
Mono Q 10/10	55	67	24.5	32	46	
Phényl-superose	10	2.2	325	14	607	
Mono Q 5/5	2.5	0.082	5750	9.2	10750	
15 Bio-sil 250	1.0	0.055	7650	8.2	14300	

c) Séquence NH<sub>2</sub>-terminale et séquences partielles internes de la precorrin-6x réductase de *Pseudomonas denitrificans* et identification du gène de structure de *Pseudomonas denitrificans* codant pour cette activité.

20 La séquence NH<sub>2</sub>-terminale de la precorrin-6x réductase de *Pseudomonas denitrificans* purifiée comme décrit ci dessus a été déterminée comme décrit précédemment. Six résidus ont été identifiés:

Ala-Gly-Ser-Leu-Phe-Asp

De même, après digestion trypsique et séparation des fragment CLHP sur 25 une colonne de phase inverse C-18, trois séquences internes ont été obtenues:

Ile-Gly-Gly-Phe-Gly-Gly-Ala-Asp-Gly-Leu

**Arg-Pro-Glu-Trp-Val-Pro-Leu-Pro-Gly-Asp-Arg**

**Val-Phe-Leu-Ala-Ile-Gly**

La séquence NH<sub>2</sub>-terminale de la protéine COBK (figure 16) correspond exactement à la séquence NH<sub>2</sub>-terminale de la precorrin-6x réductase, si ce n'est que dans la séquence présentée sur la figure 16, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminoterminale est certainement excisée *in vivo* par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). De même, les trois séquences internes correspondent aux trois séquences 60 à 69, 112 à 122, et 143 à 148 de la protéine COBK. Le poids moléculaire de la precorrin-6x réductase purifiée est estimé par électrophorèse PAGE-SDS à 31 000. La protéine COBK a un poids moléculaire déduit de sa séquence de 28 000 (figure 16). Les correspondances entre les séquences NH<sub>2</sub>-terminales, internes, et les poids moléculaires indiquent clairement que la protéine COBK correspond à la precorrin-6x réductase. Le gène cobK est le gène de structure de la precorrin-6x réductase.

d) Réaction catalysée par la precorrin-6x réductase.

La réaction enzymatique de réduction du precorrin-6x est NADPH-dépendante de façon stricte chez *P. denitrificans*. Le NADPH ne peut pas être remplacé par le NADH. Lorsque l'enzyme purifiée (ou une fraction active en cours de purification, ou même un extrait enzymatique brut) est incubée dans les conditions du dosage de l'activité, mais en l'absence de SAM et de dihydroprecorrin-6x méthylase, le produit de la réaction peut alors être purifié par CLHP dans le système décrit pour la purification du precorrin-6x (Cf exemple 6.1.4.d). Après désalage et estérification (méthanol sulfurique à 4%, 20 °C, 24h, atmosphère argon), l'ester correspondant a une masse m/z = 1008. Le produit de la réaction catalysée par la precorrin-6x reductase est donc le dihydroprecorrin-6x, appelé aussi

precorrin-6y.

6.1.8. Identification de la protéine COBQ codée par le gène CobQ.

a) Dosage de l'activité cobyric acid synthase.

5 Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été décrite. Il s'agit de la cobyric acid synthase. Cette enzyme catalyse l'amidation des fonctions acides carboxyliques périphériques en positions b, d, e, et g sur le noyau corrine (voir FIG. 19; PL. 68). Le 10 donneur de groupements NH<sub>2</sub> est la L-glutamine et chaque réaction d'amidation s'accompagne de la consommation d'une molécule d'ATP.

La fraction à doser est incubée à l'obscurité à 30°C pendant 60 min dans 250 µl de tampon 0.1 M Tris hydrochloride pH 7.5 contenant 1 mM DTT, 1 mM EDTA, 1 mM ATP, 2.5 mM MgCl<sub>2</sub>, 1 mM glutamine, 10 µM 15 acide Ado-cobyrinique di- ou pentaamide. La réaction est stoppée en ajoutant 25 µl d'une solution aqueuse 0.1 M de cyanure de potassium. Après chauffage à 80°C durant 10 min et centrifugation à 3000 x g durant 10 min, les composés formés présents dans le surnageant sont analysés par CLHP. L'unité d'activité est définie comme la quantité 20 d'enzyme nécessaire pour générer 1 nmol de fonctions amides par h dans ces conditions.

Les acides 5'-deoxy-5'-adenosyl(Ado)-cobyriniques diamide et pentaamide sont isolés de cultures de la souche SC510 en milieu PS4, en utilisant la méthode dont le principe est décrit dans l'exemple 9.

25 b) Purification de la cobyric acid synthase.

A partir du dosage décrit à l'exemple 6.1.8 a), la purification de la cobyric acid synthase de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous.

Dans une expérience typique de purification, 6 g de cellules humides 30 de SC510 Rifr dans laquelle on a introduit le plasmide pXL618 (voir

exemple 4.5.2) sont soniqués dans 15 ml de tampon Tris-HCl 0.1 M pH 7.7, DTT 1 mM, EDTA 1mM. Après centrifugation (50 000 x g durant 1 h), l'extrait est amené à 20% de glycerol (vol/vol). A 8.5 ml de l'extrait brut (203.5 mg de protéines) sont ajoutés 24 ml de tampon Tris-HCl 10 mM, DTT 1 mM, glycerol 20%. La solution est injectée sur Mono Q HR 10/10 (Pharmacia) à 2 ml/min équilibrée avec du tampon Tris-HCl 50 mM pH 7.7, DTT 1 mM, glycérol 20%. Les protéines sont éluées avec un gradient linéaire de 0.5 M NaCl et les fractions actives regroupées et amenées à 1 mM EDTA. La solution est amenée à 0.85 M en sulfate d'ammonium et injectée sur une colonne Phenyl-Superose HR 5/5 (Pharmacia), équilibrée dans le tampon Tris-HCl pH 7.7, DTT 1 mM, sulfate d'ammonium 0.85 M et les protéines sont éluées avec un gradient linéaire décroissant de 0.85 M à 0 M de sulfate d'ammonium. Les fractions sont immédiatement amenées à 20% de glycérol. La fraction active est concentrée à 2.5 ml par ultrafiltration et chromatographiée sur une colonne PD 10 (Pharmacia) équilibrée et éluée avec du tampon Tris-HCl 50 mM pH 8.3, DTT 1 mM, glycérol 20% (vol/vol). La fraction protéique est recueillie et injectée sur une Mono Q HR 5/5 équilibrée avec le même tampon et les protéines sont éluées avec un gradient linéaire de 0.5 M NaCl. La chromatographie de perméation sur gel Bio-Sil 250 (Bio-Rad) en milieu tampon Tris-HCl 50 mM pH 7.5, DTT 1 mM, glycérol 20%, NaCl 0.1 M permet finalement d'obtenir une protéine pure à plus de 97%. Elle ne présente aucune protéine contaminante en PAGE-SDS. Cette pureté est confirmée par l'unicité de la séquence NH<sub>2</sub>-terminale. Son poids moléculaire dans cette technique est de 57 000. Les différentes étapes de purification de la cobyric acid synthase avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous.

30 Tableau: Purification de la cobyric acid synthase.

	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. U /mg		Rendement <sup>a</sup>	Facteur de purification <sup>a</sup>
				A <sup>a</sup>	B <sup>b</sup>		
5	Extrait brut	8.5	203	114	/ 118	-	-
	MonoQ 10/10	8.0	35.5	388	/ 425	60	3.4
	Phényl-superose	8.0	3.23	1988	/ 2021	28	17
	MonoQ 5/5	1.0	1.20	4549	/ 4085	24	40
	Bio-Sil 250	0.75	0.88	4992	/ N.D.	19	44

10 a/ Avec l'acide Ado-cobyrinique a,c-diamide comme substrat

b/ Avec l'acide Ado-cobyrinique pentaamide comme substrat

ND = Non Déterminé

Le très haut degré de pureté de la protéine purifiée ainsi que la constance du ratio des activités d'amidation des acides cobyriniques diamide et pentaamide tout au long du procédé de purification de la protéine (voir tableau ci-dessus) indiquent sans ambiguïté qu'une seule et même protéine est responsable des quatre activités d'amidation du noyau corrine aux positions b, d, e, et g.

c) Séquence NH2-terminale de la cobyric acid synthase de Pseudomonas denitrificans et identification du gène de structure de Pseudomonas denitrificans codant pour cette activité.

La séquence NH2-terminale de la cobyric acid synthase de Pseudomonas denitrificans a été déterminée comme décrit précédemment.

Seize résidus ont été identifiés:

25 Thr-Arg-Arg-Ile-Met-Leu-Gln-Gly-Thr-Gly-Ser-Asp-Val-Gly-Lys-Ser

La séquence NH2-terminale de la protéine COBQ (figure 47) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 47, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la

méthionine aminoterminale est certainement excisée *in vivo* par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de la cobyric acid synthase purifiée est estimé par électrophorèse PAGE-SDS à 57 000. La protéine COBQ a un poids moléculaire déduit de sa séquence de 52 000 (figure 47). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires indiquent clairement que la protéine COBQ correspond à la cobyric acid synthase. Le gène cobQ est le gène de structure de la cobyric acid synthase.

10           6.1.9. Identification de la protéine COBO codée par le gène cobO.

a) dosage de l'activité cob(I)alamin adenosyltransferase (EC 2.5.1.17).

Cet exemple illustre le dosage d'une activité enzymatique directement liée à la voie de biosynthèse des cobalamines. Il s'agit de la cob(I)alamin adenosyltransferase (EC 2.5.1.17). Cette enzyme a été mise en évidence dans les cellules bactériennes (Ohta et al., 1976, Brady et al., 1962) et animales (Fenton et al. 1978). Elle a été purifiée à partir de Clostridium tetanomorphum (Vitols et al., 1966). Les fractions contenant l'activité cob(I)alamin adenosyltransferase (environ 20 unités) sont incubées de façon anaérobie à 30°C durant 15 min à l'abri de la lumière dans 1 ml de tampon Tris/HCl 0.2 M pH 8.0 en présence de 5 mM DTT, 400 μM [8-14C] ATP (2.5 μCi/μmol), 800 μM MnCl<sub>2</sub>, 50 μM hydroxocobalamine ou diaquacobinamide et 3 mg KBH4. La réaction est ensuite stoppée par chauffage à 80°C durant 10 min et, après centrifugation à 15000 x g durant 5 min, 200 μl de surnageant sont analysés par CLHP (Gimsing et al., 1986, Jacobsen et al., 1986).

L'unité d'activité enzymatique est définie comme la quantité d'enzyme nécessaire pour générer 1 nmol d'adénosylcorrinoïde par 30 min dans ces conditions.

b) Purification de l'activité cob(I)alamin adenosyltransferase.

A partir du dosage décrit à l'exemple 6.1.9 a), la purification de la cob(I)alamin adenosyltransferase de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous.

Dans une expérience typique de purification, 10 g de cellules humides de 5 la souche SC510 Rifr dans laquelle le gène cobO a été amplifié sont suspendues dans 20 ml de tampon Tris-HCl 0.2 M pH 8.0 et soniquées durant 40 min à 4°C. L'extrait brut est ensuite récupéré par centrifugation durant 1 h à 50 000 x g et dessalé sur colonnes PD10 (Pharmacia) équilibrées avec du tampon 50 mM Tris-HCl pH 8.0, 5 mM 10 DTT (tampon A). La solution protéique est ensuite fractionnée (280 mg de protéines à chaque chromatographie) sur une colonne Mono Q HR 10/10 (Pharmacia) à l'aide d'un gradient de 0 à 0.5 M KCL dans le tampon A, puis les fractions contenant l'activité sont poolées, concentrées par ultrafiltration et chromatographiées sur une colonne 15 de Phenyl-Superose HR 10/10 (Pharmacia) dans un gradient linéaire décroissant de sulfate d'ammonium (de 1,7 à 0 M), la colonne étant équilibrée dans le tampon 0.1 M Tris-HCl pH 8.0, 5 mM DTT. Pour parfaire la purification, la protéine est finalement chromatographiée, après concentration par ultrafiltration, sur une colonne Bio-Sil 250 20 (Bio-Rad) éluée avec le tampon 50 mM Tris-HCl pH 7.5, 0.1 M NaCl, 5 mM DTT.

Après cette étape, l'enzyme est pure à plus de 95%. Elle ne présente aucune protéine contaminante en PAGE-SDS. Son poids moléculaire dans cette technique est de 28 000. Ce degré de pureté est confirmé 25 par l'unicité de la séquence NH<sub>2</sub>-terminale. Les différentes étapes de purification de la cob(I)alamin adenosyltransferase, avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous pour les deux substrats suivants: diaquacobinamide (a) et hydroxocobalamine (b). Ces résultats démontrent l'absence de 30 spécificité de cette enzyme pour la nature du substrat corrinoïde. D'autre part, tous les corrinoïdes de la voie de biosynthèse entre l'acide cobyrinique diamide et la B12 ont été isolés (Blanche et al., résultats non publiés) sous leur forme native, et se sont révélés être sous forme coenzyme. Ceci démontre que le substrat naturel

de la cob(I)alamin adenosyltransferase est l'acide cobyrinique a,c-diamide.

Tableau: Purification de la cob(I)alamin adenosyltransferase.

5	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. U /mg		Rendement a	Facteur de purification <sup>a</sup>
				A	B		
10	Extrait brut	100	1400	5.4	/ 3.4	-	-
	MonoQ 10/10	90	140	34.9	/ 14.1	65	6.5
	Phényl-Superose	30	15.9	84.5	/ 49.5	18	16
	Bio-Sil 250	6.5	2.9	182.4	/ 88.7	7.0	34

c/ après dessalage sur PD10

15 c) Séquence NH2-terminale de la cob(I)alamin adenosyltransferase de Pseudomonas denitrificans et identification du gène de structure de Pseudomonas denitrificans codant pour cette activité.

La séquence NH2-terminale de la cob(I)alamin adenosyltransferase de Pseudomonas denitrificans purifiée comme décrit à l'exemple 6.1.9 b) a été déterminée, comme décrit précédemment. 13 résidus ont été 20 identifiés:

Ser-Asp-Glu-Thr-?-Val-Gly-Gly-Glu-Ala-Pro-Ala-Lys-Lys

La séquence NH2-terminale de la protéine COBO (figure 47) correspond exactement à la séquence NH2-terminale de la cob(I)alamin adenosyltransferase, si ce n'est que dans la séquence 25 présentée sur la figure 47, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminoterminal est certainement excisée in vivo par la

méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de la cob(I)alamin adenosyltransferase purifiée est estimé par électrophorèse PAGE-SDS à 28 000. La protéine COBO a un poids moléculaire déduit de sa séquence de 24 000 5 (figure 47). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires indiquent clairement que la protéine COBO correspond à la cob(I)alamin adenosyltransferase. Le gène cobO est le gène de structure de la cob(I)alamin adenosyltransferase.

6.1.10. Identification de la protéine COBN codée par le gène  
10 cobN.

a) Mise en évidence de l'activité de transformation de l'acide hydrogénobyrinique a,c-diamide en acide cobyrinique a,c-diamide.

Cet exemple illustre la mise en évidence d'une activité 15 enzymatique directement liée à la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été décrite. Il s'agit de l'activité de transformation de l'acide hydrogénobyrinique a,c-diamide en acide cobyrinique a,c-diamide.

Cette activité est mise en évidence, entre autres, par l'expérience 20 typique suivante. Un extrait brut de la souche SC510 Rifr est obtenu par sonication de 10 g de cellules humides dans 20 ml de tampon Tris/HCl 0.2 M pH 8.0, puis élimination des débris cellulaires par centrifugation durant 1 h à 50 000 x g. 1000 mg de protéines de cet extract sont incubés 1 h à 30 °C avec de l'acide hydrogénobyrinique 25 diamide marqué au carbone 14 (32 nmol; 50 µCi/µmol) dans 40 ml de tampon Tris/HCl 0.2 M pH 8.0 contenant 7 mM ATP, 200 µM CoCl<sub>2</sub>. La réaction est arrêtée par addition de 7.5 ml KH<sub>2</sub>PO<sub>4</sub> 1M et 6 ml KCN 0.3 M suivi du chauffage pendant 10 min à 80 °C. Après 30 centrifugation à 15000 x g durant 15 min, l'analyse CLHP du surnageant montre: (1) la formation durant l'incubation de 19.2 nmol d'acide cobyrinique a,c-diamide ayant la même radioactivité spécifique que l'acide hydrogénobyrinique a,c-diamide de départ et (2) la disparition

d'une quantité correspondante de ce dernier. Pour confirmer qu'il s'agit bien d'acide cobyrinique a,c-diamide, le produit est purifié par CLHP puis estérifiée dans du méthanol contenant 5% d'acide sulfurique (18h, 20°C). L'authenticité de l'ester pentaméthyle de l'acide cobyrinique a,c-diamide produit est démontré par TLC (par rapport à un échantillon de référence) et spectrométrie de masse. Notons que dans des conditions similaires d'incubation où le marquage radioactif est introduit non pas dans l'acide hydrogénobyrinique a,c-diamide mais dans le cobalt (en utilisant 10 le cobalt 57), de l'acide cobyrinique a,c-diamide marqué au cobalt 57 est biosynthétisé et les mêmes conclusions ont pu être tirées. L'acide hydrogénobyrinique a,c-diamide marqué au carbone 14 est obtenu de la manière suivante: l'acide hydrogénobyrinique est biosynthétisé in vitro en utilisant le [methyl-14C]SAM puis transformé en acide 15 hydrogénobyrinique a,c-diamide et purifié par CLHP comme décrit à l'exemple 6.1.2.

Cette étude démontre que l'insertion du cobalt a lieu au niveau de l'acide hydrogénobyrinique a,c-diamide chez *P.denitrificans*. Dans les conditions décrites, l'acide hydrogénobyrinique n'est pas substrat de la 20 chélation enzymatique par le cobalt.

b) Dosage et purification d'une protéine de la souche SC510 Rifr impliquée dans la transformation de l'acide hydrogénobyrinique a,c-diamide en acide cobyrinique a,c-diamide.

La fraction à doser (0.5 à 2 unités) est incubée pendant 60 min à 30 25 °C avec 50 µl d'extrait brut de la souche SC510 Rifr obtenu comme décrit ci-dessus, 7 mM ATP, 200 µM CoCl<sub>2</sub>, 7 µM acide hydrogénobyrinique a,c-diamide marqué au carbone 14 (50 µCi/µmol) dans 400 µl de tampon Tris/HCl 0.1 M pH 8.0. La réaction est arrêtée par addition de 75 µl KH<sub>2</sub>PO<sub>4</sub> 1 M et 60 µl KCN 0.3 M, puis chauffage 30 pendant 10 min à 80°C. Après centrifugation à 15000 x g durant 15 min, le surnageant est analysé par CLHP, pour quantifier l'acide cobyrinique a,c-diamide formé (Cf exemple 9). L'unité d'activité enzymatique est définie comme la quantité d'enzyme nécessaire pour

générer 1 nmol d'acide cobyrinique a,c-diamide par h dans ces conditions. Dans ces conditions, il apparaît que les extraits de la souche SC510 Rifr dans laquelle on a introduit le plasmide pXL1909 (voir exemple 4.5.2) présentent une activité entre 20 et 50 fois plus 5 élevée que les extraits de la souche SC510 Rifr. C'est sur cette base qu'est purifiée une protéine, seule responsable de cette amplification d'activité.

Dans une expérience typique de purification, 10 g de cellules humides de la souche SC510 Rifr dans laquelle on a introduit le plasmide 10 pXL1909 sont suspendues dans 20 ml de tampon Tris/HCl 0.2 M pH 8.0 et soniquées durant 30 min à 4°C. L'extrait brut est ensuite récupéré par centrifugation durant 1 h à 50 000 x g et dessalé sur colonnes PD10 (Pharmacia) équilibrées avec du tampon 0.1 M Tris/HCl pH 8.0 (tampon A). La solution protéique est ensuite fractionnée (213 mg de 15 protéines à chaque chromatographie) sur une colonne Mono Q HR 10/10 (Pharmacia) à l'aide d'un gradient de 0 à 0.5 M KCL dans le tampon A, puis les fractions contenant l'activité sont poolées, concentrées par ultrafiltration, dessalées sur colonnes PD10 (Pharmacia) équilibrées avec le tampon 0.1 M Tris/HCl pH 7.2 (tampon B), et 20 chromatographiées sur une colonne Mono Q HR 10/10 (Pharmacia) à l'aide d'un gradient de 0 à 0.5 M KCL dans le tampon B. Les fractions contenant l'activité sont poolées, concentrées par ultrafiltration, dessalées sur colonnes PD10 (Pharmacia) équilibrées avec le tampon B, et chromatographiées sur une colonne Mono Q HR 5/5 (Pharmacia) à l'aide d'un gradient de 0 à 0.5 M KCL dans le tampon B. 25 Pour parfaire la purification, la protéine est finalement chromatographiée sur une colonne Bio-Sil 250 (Bio-Rad) éluée avec 20 mM potassium phosphate-50 mM sodium sulfate, pH 6.8. Après cette étape, l'enzyme est pure à plus de 95%. Elle ne présente 30 aucune protéine contaminante en PAGE-SDS. Son poids moléculaire dans cette technique est de 135 000. Ce degré de pureté est confirmé par l'unicité de la séquence NH<sub>2</sub>-terminale. Les différentes étapes de purification de la protéine de la souche SC510 Rifr impliquée dans la transformation de l'acide hydrogenobyrinique a,c-diamide en acide

cobyrinique a,c-diamide, avec leur facteur de purification et leur rendement sont portées sur le tableau ci-dessous.

Tableau: Purification d'une protéine de la souche SC510 Rifr impliquée dans la transformation de l'acide hydrogenobyrinique a,c-diamide en acide cobyrinique a,c-diamide.

	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification <sup>1</sup>
10	Extrait brut	31.5	1278	0.23	-	-
15	MonoQ 10/10	44	79.2	2.4	64	10
	MonoQ 10/10	21	33.6	6.8	78	30
	MonoQ 5/5	3	6.6	16.0	36	70
15	Bio-Sil 250	1.8	5.9	16.3	33	71

c) Séquence NH<sub>2</sub>-terminale de la protéine impliquée dans la transformation de l'acide hydrogenobyrinique a,c-diamide en acide cobyrinique a,c-diamide de *Pseudomonas denitrificans* et identification du gène de structure de *Pseudomonas denitrificans* codant pour cette activité.

La séquence NH<sub>2</sub>-terminale de cette protéine purifiée comme décrit à l'exemple 6.1.10b) a été déterminée comme décrit précédemment. Six résidus ont été identifiés:

His-Leu-Leu-Leu-Ala-Gln

25 La séquence NH<sub>2</sub>-terminale de la protéine COBN (figure 47) correspond exactement à la séquence NH<sub>2</sub>-terminale de la protéine purifiée, si ce n'est que dans la séquence présentée sur la figure 47, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la méthionine aminoterminale

est certainement excisée in vivo par la méthionine aminopeptidase (Ben Bassat et Bauer, 1987). Le poids moléculaire de la protéine purifiée est estimé par électrophorèse PAGE-SDS à 135 000. La protéine COBN a un poids moléculaire déduit de sa séquence de 138 000 (figure 5 47). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires indiquent clairement que la protéine COBN correspond à la protéine impliquée dans la transformation de l'acide hydrogenobyrinique a,c-diamide en acide cobyrinique a,c-diamide. Le gène cobN est donc le gène de structure de cette protéine.

10 6.1.11. Identification de la protéine COBP codée par le gène CobP.

a) Dosage de l'activité cobinamide kinase.

Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été étudiée. Il s'agit de l'activité cobinamide kinase. Elle catalyse la phosphorylation ATP-dépendante du groupement hydroxyl du résidu (R)-1-amino-2-propanol du Ado-cobinamide pour générer le cobinamide phosphate.

La fraction à doser est incubée à l'obscurité à 30°C pendant 60 min 20 dans 500 µl de tampon 0.1 M Tris-HCl pH 8.8 contenant 1 mM EDTA, 1 mM ATP, 2.5 mM MgCl<sub>2</sub>, et 16 µM Ado-cobinamide (Blanche et al., 1989). La réaction est stoppée en ajoutant 500 µl d'une solution aqueuse 20 mM de cyanure de potassium. Après chauffage à 80°C durant 10 min et centrifugation à 5000 × g durant 10 min, le cobinamide phosphate 25 formé présent dans le surnageant est dosé par CLHP (Cf exemple 9) en utilisant le gradient linéaire simplifié suivant: de 25% à 30% de B dans A en 15 min, puis 30% à 100% de B en 12 min, et 3 min à 100% de B.

L'unité d'activité est définie comme la quantité d'enzyme nécessaire 30 pour générer 1 nmol de cobinamide phosphate à partir du cobinamide par h dans ces conditions.

b) Dosage de l'activité cobinamide phosphate guanylyltransferase.

Cet exemple illustre le dosage d'une activité enzymatique de la voie de biosynthèse des cobalamines qui n'a jusqu'à ce jour jamais été étudiée. Il s'agit de l'activité cobinamide phosphate guanylyltransferase. Elle catalyse l'addition de la partie GMP d'une molécule de GTP sur l'Ado-cobinamide phosphate, générant ainsi une molécule de GDP-cobinamide et libérant une molécule de pyrophosphate.

Cette activité est dosée dans les mêmes conditions que la cobinamide kinase si ce n'est que l'Ado-cobinamide phosphate (16 µM)(Blanche et al., 1989) et le GTP (2 mM) remplacent l'Ado-cobinamide et l'ATP respectivement, durant l'incubation.

L'unité d'activité est définie comme la quantité d'enzyme nécessaire pour générer 1 nmol de GDP-cobinamide à partir du cobinamide phosphate par h dans ces conditions.

c) Purification de la cobinamide kinase.

A partir du dosage décrit à l'exemple 6.1.11a), la purification de la cobinamide kinase de *Pseudomonas denitrificans* est réalisée comme décrit ci-dessous.

Dans une expérience typique de purification, 5 g de cellules humides de SC510 Rifr dans laquelle on a introduit le plasmide pXL623 (voir exemple 4.5.2) sont soniqués dans 20 ml de tampon Tris 0.1 M pH 7.6 (tampon A). Après centrifugation (50 000 x g durant 1 h) et dialyse 4 h contre le tampon A, 4.5 ml du rétentat sont injectés sur Mono Q HR 10/10 (Pharmacia), équilibrée avec le tampon A. Les protéines sont élues avec un gradient linéaire de 0.4 M NaCl et les fractions actives regroupées sont passées à travers une colonne de PD-10 (Pharmacia) équilibrée dans du Tris-HCl 30 mM-phosphate de potassium 5 mM-chlorure de calcium 5 µM pH 7.6 (tampon B). La solution protéique est fractionnée sur une colonne Bio-Gel HPHT (Bio-Rad), équilibrée dans le tampon B et éluee avec un gradient de 5 à 350 mM de phosphate de

potassium. Les fractions actives sont regroupées et amenées à 500 mM en sulfate d'ammonium, puis fractionnées sur une Phényl-Superose HR 5/5 (Pharmacia), éluée avec un gradient décroissant de sulfate d'ammonium. la fraction contenant l'activité est finalement 5 repurifiée sur une colonne Mono Q HR 5/5 dans le Tris-HCl à pH 7.3. Après cette étape, la protéine est pure à plus de 97%. Elle ne présente aucune protéine contaminante en PAGE-SDS. Cette pureté est confirmée par l'unicité de la séquence NH<sub>2</sub>-terminale. Son poids moléculaire dans cette technique est de 20 000. Les différentes 10 étapes de purification de la cobinamide kinase avec leur facteur de purification et leur rendement sont portées sur le tableau A.

Les fractions contenant l'activité cobinamide kinase présentent aussi l'activité cobinamide phosphate guanylyltransferase. D'autre part, comme le montrent les résultats présentés dans le tableau ci-dessus, le 15 ratio de ces deux activités reste constant dans les fractions tout au long de la purification. Enfin, la protéine purifiée présente un très haut degré de pureté dépassant les 97%. L'ensemble de ces résultats indiquent donc sans ambiguïté qu'une seule et même protéine est responsable des deux activités successives que sont la 20 cobinamide kinase et la cobinamide phosphate guanylyltransferase de la voie de biosynthèse des cobalamines chez Pseudomonas denitrificans.

d) Séquence NH<sub>2</sub>-terminale de la cobinamide kinase-cobinamide phosphate guanylyltransferase de Pseudomonas denitrificans et 25 identification du gène de structure de Pseudomonas denitrificans codant pour cette activité.

La séquence NH<sub>2</sub>-terminale de la cobinamide kinase-cobinamide phosphate guanylyltransferase de Pseudomonas denitrificans a été déterminée comme décrit précédemment. Dix résidus ont été identifiés:

30 Ser-Ser-Leu-Ser-Ala-Gly-Pro-Val-Leu-Val

La séquence NH<sub>2</sub>-terminale de la protéine COBP (figure 47) correspond exactement à cette séquence, si ce n'est que dans la séquence présentée sur la figure 47, une méthionine précède la séquence peptidique déterminée par séquençage direct. Il en résulte que la 5 méthionine aminoterminal est certainement excisée *in vivo* par la méthionine aminopeptidase (Ben Bassat et Bauér, 1987). Le poids moléculaire de la cobinamide kinase-cobinamide phosphate guanylyltransferase purifiée est estimé par électrophorèse PAGE-SDS à 20 000. La protéine COBP a un poids moléculaire déduit de sa 10 séquence de 19 500 (figure 47). Les correspondances entre les séquences NH<sub>2</sub>-terminales et les poids moléculaires indiquent clairement que la protéine COBP correspond à la cobinamide kinase-cobinamide phosphate guanylyltransferase. Le gène *cobP* est le 15 gène de structure de la cobinamide kinase-cobinamide phosphate guanylyltransferase.

#### 6.2 - Détermination des propriétés de protéines COB par mesures d'intermédiaires de biosynthèse accumulés

Cet exemple illustre comment il est possible d'attribuer à une protéine COB de Pseudomonas denitrificans une activité enzymatique. Cette activité est attribuée d'après des données obtenues concernant les intermédiaires de biosynthèse accumulés chez le ou les mutants Cob bloqués dans l'étape en question. En effet si un mutant accumule un intermédiaire de biosynthèse, il est très probable que ce mutant est bloqué dans l'étape qui a pour substrat l'intermédiaire en 20 question. 25

##### 6.2.1. Propriétés des protéines COBC et COBD

Les mutants Cob G643 (Agrobacterium tumefaciens) et G572 (Pseudomonas putida) déjà décrits dans les exemples 1 et 4 sont bloqués dans l'étape correspondant à la protéine COBC. En effet ces 30 deux mutants ne sont pas complémentés par les insertions inactivantes de transposons Tn<sub>5</sub> qui se trouvent dans le gène cobC. Les deux souches G643 et G572, ainsi que les souches parentes non mutées [C58-C9 Rif<sup>r</sup> et KT 2440 Rif<sup>r</sup> (Cameron et al., 1989)] ont été cultivées en milieu

PS4' pour les *A.tumefaciens*, et PS4'' pour les *P.putida* (PS4' et PS4'' correspondent à du milieu PS4 contenant respectivement 100 et 1000 fois moins de cobalt que le PS4 décrit précédemment) 3 jours comme cela est décrit ci-dessus. Du  $^{57}\text{CoCl}_2$  a été ajouté aux cultures (2,5 5  $\mu\text{Ci}/0,1 \mu\text{m}$  pour une culture de 25 ml). Les corrinoïdes intracellulaires ont été isolés sous leur forme native et identifiés par leur comportement CLHP. Les souches parentes n'accumulent pas de corrinoïdes autres que du coenzyme  $B_{12}$ . Les deux mutants, G643 et G572, accumulent de l'acide cobyrique adénosylé dans des proportions 10 respectives de 11 % et 6 %. Ces proportions en % sont calculées par rapport au niveau de coenzyme  $B_{12}$  synthétisé par la souche parente. Outre l'acide cobyrique, le mutant G643 accumule de l'acide cobyrinique pentaamide dans une proportion de 2 %; l'acide cobyrinique pentaamide est l'intermédiaire qui précède l'acide cobyrique. L'étude 15 de ces mutants fait ressortir qu'ils sont bloqués après l'acide cobyrique. Tous ces mutants Cob sont soit bloqués entre l'uro'gen III et le cobinamide, soit entre le cobinamide et les cobalamines. Les mutants G643 et G572 sont bloqués entre l'uro'gen III et le cobinamide. Or si ces mutants sont bloqués avant le cobinamide et 20 accumulent tous les deux de l'acide cobyrique, les protéines pour lesquelles ils codent ne peuvent intervenir que dans l'étape enzymatique (appelée cobinamide synthase) qui catalyse l'amidation de l'acide cobyrique par un résidu aminopropanol pour donner le cobinamide; elles peuvent aussi éventuellement intervenir dans la 25 synthèse du substrat de la réaction qui apporte l'aminopropanol, si ce n'est l'aminopropanol lui-même. Le gène cobC code pour une protéine qui est soit la cobinamide synthase, soit une de ses sous-unités.

Le mutant Cob G634 d'*Agrobacterium tumefaciens* qui est bloqué dans l'étape correspondant au gène cobD a été analysé de la 30 même manière. Ce mutant n'est pas complémenté par les insertions inactivationnelles dans le gène cobD (exemple 4.1). Le seul corrinoïde intracellulaire trouvé chez ce mutant est de l'acide cobyrique adénosylé. Comme les mutants précédents, ce mutant code pour une protéine intervenant dans la transformation de l'acide cobyrique en 35 cobinamide ou bien éventuellement dans la synthèse de l'autre substrat

de la réaction.

Ces deux gènes différents (cobC et cobD) codent pour deux protéines qui interviennent dans la même étape.

#### 6.2.2. Propriétés des protéines COBF à COBM

5 Les mutants d'Agrobacterium tumefaciens déjà décrits dont on connaît, d'après l'étude décrite à l'exemple 4.2, dans quels gènes chacun d'entre eux est bloqué, ont été étudiés. Ce sont les mutants : G612 (cobF), G615 (cobG), G616 (cobH), G613 (cobI), G611 (cobJ), G620 (cobK), G638 (cobL) et G609 (cobM) ; nous avons figuré entre 10 parenthèses le gène de Pseudomonas denitrificans responsable de la complémentation de ces mutants (exemple 5) qui correspond donc au gène muté chez ce mutant. Ces mutants ont été cultivés en milieu PS4 comme cela est décrit précédemment avec du cobalt marqué. Après quatre jours 15 d'incubation, les mutants ont été analysés pour leur contenu intracellulaire en corrinoïdes et descobaltocorrinoïdes (Cf exemples 6.1.2 et 9).

Tableau : Intermédiaires accumulés par des mutants d'Agrobacterium tumefaciens bloqués dans les gènes du fragment 8,7 kb de Pseudomonas denitrificans

'p Souches	Descobaltocorrinoïdes intracellulaires en % <sup>1</sup>			Corrinoides intracellulaires en % des cobalamines	Gène muté
	HAB	HABM	HABD		
5	C58-C9 <sup>*</sup>	100	100	100	-
	G612	< 5	< 5	64	<u>cobF</u>
	G615	< 5	< 5	84	<u>cobG</u>
	G616	35	< 10	< 10	<u>cobH</u>
	G613	< 5	< 5	57	<u>cobI</u>
	G611	< 5	< 5	65	<u>cobJ</u>
	G620	12	< 5	< 10	<u>cobK</u>
	G638	< 5	< 5	47	<u>cobL</u>
	G609	< 5	< 5	33	<u>cobM</u>

HAB : acide hydrogénobyrrinique

HABM : acide hydrogénobyrrinique monoamide

15 HABD : acide hydrogénobyrrinique diamide

\* en fait il s'agit de la souche C58-C9 Rif<sup>r</sup>Nal<sup>r</sup> déjà décrite (Cameron et al., 1989)

<sup>1</sup>les valeurs sont exprimées en % des mêmes intermédiaires accumulés chez la souche parente non mutée C58-C9 Rif<sup>r</sup>Nal<sup>r</sup>.

20 Ces résultats montrent que tous les mutants n'accumulent aucun corrinoïde (à l'exception du mutant inactivé dans le gène cobF, G612, qui lui, accumule du cobinamide mais à un niveau faible équivalent à 2,2 % des cobalamines synthétisées par la souche non mutée). Toutefois, certains mutants (G612, G615 et G616) ont des niveaux de cobalamines qui représentent plus de 10 % du niveau de cobalamines de la souche parente. Il est probable que tous ces mutants sont bloqués au moins avant l'acide cobyrinique diamide. Tous ces mutants accumulent de l'acide hydrogénobyrrinique et de l'acide hydrogénobyrrinique diamide en quantités inférieures à la souche non

mutée; ils sont donc très probablement bloqués avant l'acide hydrogénobyrinique. Nous pouvons conclure que tous les gènes cobF à cobG codent pour des protéines qui interviennent avant l'acide hydrogénobyrinique. On sait que le mutant G613 est muté dans le gène 5 cobi qui code pour la SP<sub>2</sub>MT, intervenant bien avant l'acide hydrogénobyrinique. Pour ce mutant, les résultats du présent exemple, concernant l'accumulation d'intermédiaires, sont en parfait accord avec l'étape inactivée chez ce mutant, à savoir que ce mutant n'accumule aucun intermédiaire après l'acide hydrogénobyrinique à un 10 niveau supérieur à ce que l'on observe avec la souche non mutée. Ce résultat est pour les gènes cobF, cobJ, cobL et cobM cohérent avec ceux de l'exemple 6.4 où il est proposé que ces gènes codent pour des protéines qui catalysent des transferts de méthyl SAM dépendants et donc qui interviennent avant l'acide hydrogénobyrinique. A l'exception 15 de cobi qui est le gène de structure de la SP<sub>2</sub>MT ces gènes interviennent après le précorrine-3. En effet comme ce ne sont ni les gènes de structure de la SUMT et de la SP<sub>2</sub>MT ils interviennent forcément après, c'est-à-dire après le précorrine-3 (tous les gènes cob décrits dans la présente invention interviennent entre l'uro'gen 20 III et les cobalamines). Ces gènes cobF à cobH et cobj à cobM codent pour des enzymes qui interviennent entre le précorrine-3 et l'acide hydrogénobyrinique.

#### 6.2.3. Propriétés des protéines COBS et COBT

Le mutant G2035 décrit dans les exemples 1 et 4.3 est 25 bloqué dans l'étape correspondant à la protéine COBS. Le mutant G2037 décrit dans l'exemple 1 est bloqué dans l'étape correspondant à la protéine COBT. Ces souches ainsi que la souche parente (*Agrobacterium tumefaciens* C58C9Rifr) sont cultivées en milieu PS4' (il s'agit du milieu PS4 où la concentration en chlorure de cobalt est 100 fois plus 30 faible que dans le milieu PS4) en présence de cobalt radioactif <sup>57</sup>CoCl<sub>2</sub> pendant 3 jours, leur contenu intracellulaire en descobaltocorrinoïdes est analysé ainsi que le contenu en corrinoïdes, comme cela a déjà été décrit ci-dessus (Cf exemple 6.2.2). Les souches

G2035 et G2037 n'accumulent pas de corrinoïdes et de larges concentrations (supérieures à celles observées avec la souche parente) en acide hydrogénobyrinique et acide hydrogénobyrinique mono et diamide ne sont présentes qu'avec la souche G2035. Ce mutant est 5 probablement bloqué dans une étape située après l'acide hydrogénobyrinique diamide et avant l'acide cobyrinique diamide. Par conséquent le gène cobS coderait pour une des enzymes impliquées dans la transformation de l'acide hydrogénobyrinique diamide en acide cobyrinique diamide; cette protéine peut donc intervenir soit dans 10 l'insertion du cobalt, soit dans la réduction du cobalt de l'acide cobyrinique a,c-diamide non adénosylé. Par contre, le mutant G2037 serait bloqué dans une étape située en amont de l'acide hydrogénobyrinique. Le gène cobT coderait pour une protéine impliquée dans une étape enzymatique en amont de l'acide hydrogénobyrinique et 15 en aval du précorrine-3 (d'autres gènes de structure codant pour les enzymes impliqués en aval du précorrine-3 ont déjà été identifiés). Une autre possibilité pour la protéine COBT est qu'elle intervienne, comme cela est proposé à l'exemple 5, comme protéine liant le cobalt, et/ou comme protéine qui interagit avec d'autre(s) protéine(s) par sa 20 partie acide.

#### 6.2.4. Propriétés de la protéine COBV

Les mutants G2039 et G2040 décrits dans les exemples 1 et 4.4 sont bloqués dans l'étape correspondant à la protéine COBV. Ces souches ainsi que la souche parente sont cultivées en milieu PS4' pendant 3 jours en présence de cobalt radioactif  $^{57}\text{CoCl}_2$ , puis leur contenu intracellulaire en descobaltocorrinoïdes est analysé et le contenu en corrinoïdes est déterminé, comme cela est décrit dans l'exemple 9. Les souches G2039 et G2040 accumulent de l'acide cobyrique, du cobinamide, du cobinamide phophate et du GDP-cobinamide. 25 Ces mutants sont probablement bloqués dans une étape enzymatique en aval de la GDP-cobinamide. Le gène cobV coderait pour une enzyme impliquée dans la transformation de la GDP-cobinamide en cobalamine, voir figure 5. Ce résultat est en parfait accord avec l'activité 30

cobalamin (5'-phosphate) synthase de la protéine COBV qui possède l'Ado-GDP-cobinamide comme substrat.

6.3 - Détermination de l'activité de protéines COB par études d'affinité vis-à-vis du SAM

5 Cet exemple illustre comment il est possible, à partir de protéines COB purifiées de Pseudomonas denitrificans, de mettre en évidence in vitro une activité de fixation du SAM. Si une protéine COB possède une telle activité, cela signifie que cette protéine COB est une méthyltransférase de la voie et qu'elle intervient dans un des  
10 transferts des huit groupements méthyl qui se produisent entre l'uro'gen III et l'acide cobyrinique.

6.3.1. Test d'affinité pour le SAM sur une protéine purifiée

15 Le test repose sur le principe suivant lequel les méthyltransférases de la voie de biosynthèse des cobalamines ont certainement un site de fixation du SAM. Ce site doit être mis en évidence par une affinité du SAM plus élevée que pour toute protéine qui ne fixe pas le SAM spécifiquement. Après incubation de la protéine à étudier en présence d'un excès de SAM radioactif, celle-ci est  
20 séparée du SAM libre par une chromatographie de perméation de gel. La radioactivité retrouvée dans la fraction du poids moléculaire de la protéine correspond au SAM fixé durant l'incubation. La chromatographie est effectuée à 2°C pour limiter au maximum la libération du SAM lié durant la séparation.

25 La protéine (environ 10 µg) est incubée durant 10 minutes à 30°C dans 200 µl de Tris-HCl 0,1 M pH 7,7 avec 5 nmoles de [méthyl-<sup>3</sup>H]-SAM (1 µCi). Après incubation, 100 µl du mélange sont immédiatement injectés sur une colonne de TSK-125 (Bio-Rad) élue à 1 ml/minute avec le mélange de sulfate de sodium 50 mM/dihydrogénophosphate de sodium 20 mM pH 6,8 préconisé par le distributeur de cette colonne. Des fractions de 0,5 ml sont collectées et comptées en scintillation liquide. Les temps de rétention de la protéine et du SAM sont obtenus directement après l'enregistrement

de l'absorbance de l'éluat à 280 nm.

6.3.2. Etude in vitro de la fixation du SAM sur les protéines COBA et COBF de Pseudomonas denitrificans

a) Purification des protéines COBF et COBA

La protéine COBF de Pseudomonas denitrificans est purifiée comme cela est décrit ci-après. Dans une expérience typique de purification, 5 g de cellules humides de la souche SC150 Rif<sup>r</sup> dans laquelle a été introduit le plasmide pXL1546 (voir ex.7.3) obtenues après culture en milieu PS4 sont resuspendues dans 30 ml de Tris-HCl 0,1 M pH 7,7 et soniquées durant 15 minutes à 4°C. L'extrait brut est ensuite récupéré par centrifugation 1 heure à 50 000 g et le surnageant est passé sur une colonne de DEAE-Sephadex (1 ml de gel) pour éliminer les composés tétrapyrroliques présents. 10 mg de protéines (0,7 ml) de cet extrait sont ensuite injectés sur une colonne de MonoQ HR 5/5 équilibrée avec le même tampon. Les protéines sont éluées avec un gradient linéaire de KCl (0 à 0,25 M). La protéine COBF est éluée avec 0,20 M de KCl. Elle est diluée deux fois avec du Tris-HCl 0,1 M pH 7,7 et purifiée une seconde fois sur une MonoQ HR 5/5. L'électrophorèse en PAGE-SDS avec révélation au bleu de Comassie est utilisée pour visualiser la protéine. Cette technique montre d'autre part que COBF est d'une pureté de 95 % environ après cette étape de purification. La séquence NH<sub>2</sub>-terminale de la protéine purifiée a été réalisée comme cela est décrit précédemment. Deux séquences NH<sub>2</sub>-terminales apparaissent en même temps sur chaque cycle de dégradation; ce sont les séquences suivantes dans les proportions indiquées :

Séquence 1 (34 % d'abondance)

1	2	3	4	5	6	7	8	9	10	11
Ala	Glu	Ala	Gly	Met	Arg	Lys	Ile	Leu	Ile	Ile

Séquence 2 (66 % d'abondance)

1	2	3	4	5	6	7	8	9	10	11
Met	Arg	Lys	Ile	Leu	Ile	Ile	Gly	Ile	Gly	Ser

La séquence 1 correspond à la séquence NH<sub>2</sub>-terminale de la protéine COBF qui est donnée sur la figure 16, si ce n'est que la

méthionine aminoterminal est excisée suivant des règles déjà énoncées (Hirel et al., 1989) par la méthionine aminopeptidase (Ben Bassat et Bauer, 1989). La séquence 2, présente en quantité la plus importante, correspond à la même protéine mais dont l'initiation de traduction se 5 serait faite non pas au codon ATG d'initiation de la traduction que nous avions supposé, mais à celui situé 5 codons en aval sur la phase codante (figure 16). En effet les acides aminés de cette séquence sont exactement ceux que l'on trouve sur la séquence de la protéine COBF à partir de la deuxième méthionine (acide aminé n° 6) de cette séquence 10 (figure 16). Dans ce cas la méthionine aminoterminal n'est pas excisée ce qui confirme les règles déjà énoncées (Hirel et al., 1989). Il y a chez la souche SC510 Rif<sup>r</sup> portant le plasmide pXL1546, deux initiations de traduction, d'une part celle correspondant au codon méthionine positionné à la bonne distance, sur notre construction, de 15 la séquence de Shine et Delgarno et d'autre part celle qui se fait au deuxième codon méthionine se trouvant sur la séquence du gène cobF présenté à la figure 16. De cela, il ressort que probablement la protéine COBF commence non pas à la méthionine indiquée sur la figure 16 mais à celle se trouvant 5 acides aminés plus loin.

20 De toute manière, ce résultat montre qu'il s'agit bien de la protéine COBF qui est exprimée et que celle-ci est exprimée avec une forme allongée de 4 acides aminés. Lors de la purification les deux formes protéiques sont purifiées. Dans cet exemple, nous appellerons protéine COBF purifiée le mélange de ces deux protéines 25 purifiées.

La protéine COBA de Pseudomonas denitrificans est purifiée comme cela est décrit précédemment (Blanche et al., 1989).

b) Fixation du SAM

La fixation du SAM sur ces deux protéines est étudiée 30 comme cela est décrit précédemment à l'exemple 6.3.1.a). La sérum albumine bovine et la protéine COBH purifiée sont utilisées comme contrôles négatifs. Pour les protéines COBA et COBF on observe un pic de radioactivité à la sortie de la colonne de TSK-125 au temps de sortie de ces protéines (figure 20). Dans ce test, la protéine COBI

présente la même propriété de fixation du SAM. A l'opposé, il n'y a pas de tels pics de radioactivité avec la BSA et la protéine COBH. Ce test met en évidence la fixation in vitro du SAM sur les protéines COBA, COBI et COBF. Ces résultats montrent que COBA, COBI et COBF sont des SAM-méthyltransférases. Ce résultat est tout à fait en accord avec les activités de COBA et COBI puisqu'il s'agit respectivement de la SUMT et de la SP<sub>2</sub>MT de Pseudomonas denitrificans. La protéine COBF est donc probablement une SAM méthyltransférase de la voie de biosynthèse des cobalamines. Ce test confirme que COBF est une méthyltransférase.

10 6.4 - Détermination de l'activité de protéines COB par études d'homologies de séquences

Cet exemple illustre comment par des comparaisons entre les séquences de diverses protéines COB de Pseudomonas denitrificans il est possible de trouver les protéines COB qui sont des SAM-méthyltransférases de la voie de biosynthèse des cobalamines.

Les protéines COBI et COBA sont toutes les deux des SAM méthyltransférases de la voie de biosynthèse. Ces deux protéines ont été comparées suivant le programme de Kanehisa, 1984. Cette comparaison fait ressortir trois régions de forte homologie (figure 21). Dans chacune de ces régions il y a plus de 45 % d'homologie stricte entre les deux protéines. Trois régions de forte homologie entre COBA et CYSG sont également présentées (figure 22); ce sont les mêmes régions de COBA qui présentent une forte homologie avec COBI. Ces régions de fortes homologies entre COBA, CYSG et COBI présentent de l'homologie avec d'autres protéines COB. Il s'agit des protéines COBF, COBJ, COBL et COBM (figure 23). En ce qui concerne la région 1, les protéines COBF, COBL et COBM présentent des homologies significatives par rapport à toutes les protéines de Genpro, qui est une extraction protéique de Genbank (version 59) augmentée des parties codantes putatives supérieures à 200 acides aminés, suivant le programme de Kanehisa (1984). En ce qui concerne la région 2, les protéines COBJ, COBL et COBM présentent des homologies significatives par rapport à toutes les protéines de Genpro (version 59). En ce qui concerne la troisième région d'homologie, COBJ, COBL et COBM

présentent des homologies significatives par rapport à toutes les protéines de Genpro (version 59). Les comparaisons de séquences permettent donc de mettre en évidence que quatre protéines COBF, COBJ, COBL et COBM présentent des homologies significatives avec les régions conservées des séquences de trois types de méthyltransférases COBA, COBI et COBF. Les protéines COBG, COBH et COBK ne présentent pas d'homologies significatives avec les régions conservées des méthylases. La protéine COBF ne présente d'homologie significative avec les autres protéines que dans la région 1. Ces homologies doivent correspondre probablement au fait que toutes ces protéines sont des méthyltransférases. Ce résultat recoupe les données biochimiques décrites sur COBF concernant la capacité qu'a cette protéine à lier le SAM in vitro (exemple 6.3). Ces homologies, d'une part, permettent de confirmer que COBF est une SAM méthyltransférase de la voie de biosynthèse des cobalamines et d'autre part mettent en évidence que COBJ, COBL et COBM pourraient être des SAM méthyltransférases de la voie de biosynthèse des cobalamines. Ces résultats montrent également l'homologie qui existe entre les protéines COB de P.denitrificans et les protéines isofonctionnelles d'autres microorganismes.

20 EXEMPLE 6(B) - Purification et clonage du gène de structure de la SUMT de Methanobacterium ivanovii.

Cet exemple illustre comment il possible d'obtenir, dans d'autres microorganismes, des enzymes COB et des gènes cob correspondants à ceux identifiés chez P. denitrificans.

25 6(B).1. Purification de la SUMT de Methanobacterium ivanovii.

Cet exemple décrit la purification de la SUMT de Methanobacterium ivanovii et l'étude de ses propriétés catalytiques. La souche de Methanobacterium ivanovii DSM2611 est cultivée comme cela est décrit (Souillard et al., 1988). 12 g de cellules humides sont obtenus. Celles ci sont resuspendues dans 80 ml de tampon Tris/HCl 0.1 M pH 7.6 contenant 5 mM DTT et 1 mM EDTA et soniquées pendant 1h30 à 4 °C puis centrifugées durant 1h à 50 000 g. L'extrait est ensuite débarassé de composés tétrapyrroliques libres par passage à travers une petite

colonne de DEAE-Sephadex A25 montée dans le même tampon. Les protéines précipitant entre 55 et 75% de saturation en sulfate d'ammonium sont solubilisées dans un tampon Tris/HCl 0.1M pH 7.5, 0.5 mM DTT, 1.7 M sulfate d'ammonium et injectées sur une colonne de Phényl-Superose HR 5 10/10 (Pharmacia France/SA), éluée avec un gradient décroissant (de 1.7 M à 0 M en sulfate d'ammonium). Les fractions actives sont passées sur une colonne de Séphadex G-25 équilibrée avec du tampon Tris/HCl 0.1 M, pH 7.5, 0.5 mM DTT, glycérol 25% (tampon A) puis injectées sur une Mono Q HR 5/5 (Pharmacia France SA) équilibrée avec du tampon A et éluée avec 10 un gradient de KCl de 0 à 0.3 M; cette étape est répétée une seconde fois dans les mêmes conditions. Une chromatographie de perméation de gel sur Bio-Sil TSK-250 (BioRad France SA) de la fraction active de l'étape précédente permet d'obtenir une protéine homogène en PAGE-SDS et en RP-CLHP (C-18 μBondapak). Les différentes étapes de purification avec 15 leur rendement ainsi que leur facteur de purification sont décrites sur le tableau ci-dessous.

Comme il est montré dans ce tableau, le facteur de purification total est de plus de 4 500. Des propriétés de l'enzyme pure ont été étudiées selon des méthodes déjà décrites (Blanche et al., 1989). Cette 20 enzyme a bien une activité SUMT, c'est à dire, qu'elle catalyse bien le transfert de deux groupements méthyls, SAM-dépendant, en C-2 et en C-7 de l'uro'gen III. Le poids moléculaire de l'enzyme estimé par perméation de gel est de 60,000 +/- 1,500 tandis que par PAGE-SDS il est de 29,000 ce qui montre clairement qu'il s'agit d'une enzyme homodimérique. Dans 25 des conditions déjà décrites (Blanche et al., 1989) l'enzyme à un Km pour l'uro'gen III de 52 +/- 8 nM. De plus, cette enzyme ne présente pas d'inhibition par son substrat à des concentrations inférieures à 20 μM, alors que la SUMT de Pseudomonas denitrificans présente une inhibition par l'uro'gen III à une concentration supérieure à 2 μM (Blanche et al., 30 1989).

Tableau : Purification de la SUMT de *M. ivanovii*.

	Etape de purification	vol (ml)	Protéines (mg)	Activité spé. (u/mg de protéines)	Rendement	Facteur de purification <sup>1</sup>
5	Extrait brut	92	731	0.337		
	55-75% SA	7.1	153	1.215	76	3.6
	Phényl-superose	9.5	8.34	15.35	52	46
	MonoQ 5/5	1.0	0.252	422	43	1252
	Bio-Sil TSK	1.0	0.061	1537	38	4561

1/ calculé d'après le rendement en protéines.

La Vmax de la SUMT de *M. ivanovii* a été déterminée. Celle-ci est de 1537 U/mg de protéines. Cette valeur est supérieure à celle trouvée pour la SUMT de *P. denitrificans* déjà déterminée dans les conditions optimales de la réaction (compte tenu de son inhibition par l'uro'gen III), 489 U/mg de protéines (Blanche et al., 1989).

#### 6(B).2. Clonage du gène de structure de la SUMT de *M. ivanovii* chez *E. coli*.

6(B).2.1. Clonage d'un fragment interne au gène de structure de la SUMT de *M. ivanovii*. Pour ce faire, on procède de la manière suivante : 200 picomoles de la SUMT de *M. ivanovii* sont utilisées pour le séquençage NH2-terminal de la protéine comme cela est décrit précédemment. En outre, un fragment peptidique obtenu par une digestion trypsique de la protéine est lui aussi soumis à un séquençage de sa partie NH2-terminale. Les séquences obtenues sont présentées sur la figure 48. Les oligonucléotides sens et antisens respectivement 946, 923 et 947 (voir figure 48) sont synthétisés comme cela est décrit précédemment; ces oligonucléotides contiennent à leur extrémité 5' un site de restriction qui est soit EcoRI pour les oligonucléotides sens soit HindIII pour l'oligonucléotide antisens. Ces oligonucléotides sont utilisés pour une expérience d'amplification enzymatique de l'ADN (Saiki et al., 1988), comme cela est schématisé sur la figure 48.B.

L'ADN génomique de *M. ivanovii* est préparé de la manière suivante: 0.4 g de cellules de *M. ivanovii* (DSM 2611) sont lavées avec une solution 0.15 M NaCl. Les cellules sont ensuite incubées dans 4 ml d'une solution de sucrose 25 %, Tris/HCl 50 mM pH 8, lysozyme 40 mg puis 5 2 à 3 h à 50 °C après addition de 40 mg de protéinase K et de 5 ml d'une solution SDS 0.2 %, EDTA 0.1 M pH 8. L'ADN est ensuite extrait au phénol-chloroforme (50 % - 50 %) 2 fois, puis 2 fois au chloroforme et ensuite précipité par l'isopropanol et repris dans 3 ml de TNE (Tris/HCl 10 mM pH 8, EDTA 1 mM, Na-Cl 100 mM).

10 L'Amplification enzymatique de l'ADN de *M. ivanovii* est effectuée suivant le protocole de Saiki et al., 1988, dans un volume de 0.1 ml avec 600 ng d'ADN génomique de *M. ivanovii*, en utilisant les amorces 946 et 947 (réaction 1) ou 923 et 947 (réaction 2). Le tampon utilisé pour cette réaction est MgCl<sub>2</sub> 1 mM, KCl 50 mM, gélatine 0.001 % 15 et chaque dNTP à une concentration de 0.2 mM; pour chaque réaction d'amplification, 10 ng de chaque oligonucléotide sont utilisés ainsi que 2.5 unités de Taq DNA polymérase (Cetus Corporation). L'amplification est réalisée sur 30 cycles dans le Perkin-Elmer Cetus DNA Amplification system; au cours de chaque cycle, l'ADN est dénaturé 1 min à 95 °C, les 20 amorces oligonucléotidiques sont hybridées à l'ADN simple brin 2 min à 38 °C et les néobrins sont polymérisés pendant 3 min à 72 °C. Les produits d'amplification sont ensuite extraits au chloroforme et subissent ensuite une précipitation éthanolique; ils peuvent ensuite être visualisés après migration sur gel d'acrylamide puis être digérés 25 par les enzymes de restriction telles que EcoRI et HindIII.

Dans le cas de la réaction 1, deux fragments sont observés : à 30 615 pb ainsi qu'à 240 pb. En ce qui concerne la réaction 2, deux fragments sont aussi observés : à 630 et 170 pb. La totalité du produit d'une réaction d'amplification enzymatique entre les oligonucléotides 35 946-947 est séparée par migration sur gel d'acrylamide; le fragment de 615 pb est purifié comme cela est décrit précédemment. Ce fragment est ensuite digéré par EcoRI et HindIII afin de rendre les extrémités du fragment cohésives. Ce fragment est ensuite ligaturé avec de l'ADN de la forme réplicative du phage M13mp19. La ligature est transformée dans *E. coli* TG1. Six clones recombinants, contenant un insert de 615 pb, sont

analysés par séquençage avec le primer universel -20 (Pharmacia SA, France). Comme il est montré sur la figure 49, lorsque l'on séquence l'ADN simple brin des phages recombinants qui contiennent l'insert de 615 pb, il doit être observé, en aval du site EcoRI, une séquence non 5 dégénérée correspondant à celle de l'oligonucléotide 946 suivie, dans la même phase, par une séquence qui code pour les acides aminés LITLKAVNVLK?ADVVL (? signifie qu'à cette position, le résidu n'a pas pu être déterminé); cette séquence correspond à celle qui, dans la séquence NH<sub>2</sub>-terminale de la SUMT, suit les acides aminés correspondant à 10 l'oligonucléotide 946 (voir figure 48). Pour deux clones, il a effectivement été observé, après le site EcoRI, une séquence pouvant coder pour la région NH<sub>2</sub>-terminale de la SUMT de *Methanobacterium ivanovii*, celle-ci commençant par l'enchaînement Pro-Gly-Asp-Pro-Glu-Leu qui sont les acides aminés codée par une séquence contenant 15 l'oligonucléotide 946. Cette observation montre que ces deux formes répliquatives recombinantes contiennent un insert qui correspond à un fragment interne au gène de structure de la SUMT de *Methanobacterium ivanovii*. La forme répliquative portant ce fragment interne au gène de structure de *M. ivanovii* est appelée pG10.

20 6(B).2.2. Clonage du gène de structure de la SUMT de *Methanobacterium ivanovii*.

L'ADN génomique de *Methanobacterium ivanovii* est digéré par plusieurs enzymes de restriction (digestions simples ou doubles). Après digestion, les fragments sont séparés par électrophorèse sur gel 25 d'agarose, puis ils sont transférés sur une membrane de nylon comme cela est décrit précédemment. Après dénaturation des fragments ainsi transférés et préhybridation, on procède à une hybridation avec la forme répliquative pG10 comme sonde marquée au <sup>32</sup>P comme cela est décrit précédemment. Il est ainsi trouvé qu'un fragment issu d'une digestion 30 EcoRI-BglII de 3.2 kb de *Methanobacterium ivanovii* hybride avec la sonde (voir Figure 50). 40 µg d'ADN génomique de *M. ivanovii* sont ensuite digérés par EcoRI et BglII puis séparés par migration sur gel d'agarose. Les fragments ayant une taille comprise entre 3 et 3.5 kb sont électroélusés comme cela est décrit précédemment. Les fragments ainsi

purifiés sont mis à ligaturer avec le vecteur pBKS+ (Stratagene Cloning Systems, La Jolla) digéré par BamHI-EcoRI. La ligature est transformée dans *E. coli* DH5 $\alpha$  (Gibco BRL). Les transformants sont sélectionnés sur milieu LB supplémenté avec de l'ampicilline et du X-gal. 800 colonies blanches sont repiquées sur filtre; après croissance puis lyse des bactéries on procède à une hybridation sur colonies selon la technique de Grünstein et Hogness (1975). La sonde utilisée est la forme réplicative pG10 marquée au 32P. Un seul clone positif après ce test d'hybridation avec la sonde est trouvé. L'ADN plasmidique de ce clone est appelé pXL1809 (voir figure 56). Une digestion de cet ADN par EcoRI-XbaI permet de visualiser, comme attendu, un insert de 3.2 kb. Le plasmide pXL1809 est séquencé sur les 2 brins par la technique de Chen et Seeburg (1985). Une séquence de 955 bases est obtenue (figure 51). Une analyse des phases ouvertes nous conduit à identifier une phase ouverte de la base 34 (ATG) à la base 729 (TGA). Cette phase ouverte code pour une protéine dont la séquence est présentée sur la figure 52. Cette protéine à un poids moléculaire de 24 900 (voir figure 53) ce qui est proche du poids moléculaire de la protéine purifiée à partir de *M. ivanovii*. La séquence NH<sub>2</sub>-terminale de cette protéine est exactement celle déterminée pour la SUMT de *M. ivanovii* purifiée (voir figure 48 et figure 52). Ces observations établissent sans ambiguïté que le gène cloné et séquencé est bien le gène de structure de la SUMT de *M. ivanovii*. Comme cette activité est supposée intervenir dans la biosynthèse des corrinoïdes chez toutes les bactéries, ce gène est désigné par gène corA et la protéine codée par ce même gène protéine CORA. Le profil d'hydrophobicité de la protéine CORA de *M. ivanovii*, réalisé à partir du programme de Hopp et Woods (1981) montre qu'il s'agit, comme cela est attendu, d'une protéine hydrophile, comme cela est présenté sur la figure 54. La protéine CORA de *M. ivanovii* montre un degré d'homologie stricte de plus de 40 % vis à vis de COBA de *P. denitrificans* (figure 53). Cette homologie s'étend sur presque la totalité des deux protéines puisqu'elle concerne les résidus 3 à 227 de CORA de *M. ivanovii* et les résidus 17 à 251 de COBA de *P. denitrificans*. Cette homologie reflète les homologies structurales qu'il existe entre deux protéines catalysant la même réaction. Ce sont les mêmes régions

qui sont les plus conservées entre CORA et COBA de *P. denitrificans* que celle qui sont conservées entre COBA de *P. denitrificans* et CYSG d'*E. coli* (figures 22).

#### EXEMPLE 7 - Expression de protéines COB

##### 7.1 - Expression chez *Pseudomonas denitrificans*

5 Cet exemple illustre que l'amplification d'un gène de structure d'une protéine COB de *Pseudomonas denitrificans* chez *Pseudomonas denitrificans* conduit à l'amplification de l'activité de la protéine COB.

###### 7.1.1 - Expression de la protéine COBA

10 Le plasmide pXL557 correspond au plasmide pXL59 dans lequel le fragment BglIII-EcoRV (respectivement aux positions 80 et 2394 sur la séquence de la figure 7) de 2,4 kb du fragment de 5,4 kb a été cloné. Ce fragment contient les gènes cobA et cobE.

15 Le plasmide pXL545 contient uniquement le gène cobE. Sa construction a été décrite dans l'exemple 4.1.

Ces deux plasmides ont été introduits par transfert conjugatif chez SC510 Rif<sup>r</sup>. Les souches SC510 Rif<sup>r</sup>, SC510 Rif<sup>r</sup> pXL59, SC510 Rif<sup>r</sup> pXL557 et SC510 Rif<sup>r</sup> pXL545 ont été cultivées en milieu PS4. A 4 jours, les cultures ont été arrêtées et les activités SUMT 20 ont été dosées suivant un protocole standard déjà décrit (F. Blanche et al., 1989). Les activités sont portées ci-dessous.

Tableau : Activité SUMT de SC510 Rif<sup>r</sup> et de quelques unes de ses dérivées

Souche	SUMT dosée nmole/h/mg de protéines
SC510 Rif <sup>r</sup>	0.05
SC510 Rif <sup>r</sup> pXL59	0.04
SC510 Rif <sup>r</sup> pXL557	2.10
SC510 Rif <sup>r</sup> pXL545	0.05

Il ressort clairement de ces résultats que seul le plasmide pXL557 entraîne chez SC510 Rif<sup>r</sup> une nette augmentation de l'activité SUMT (un facteur 50). Cette augmentation résulte de l'amplification de cobA et non de cobE puisque le plasmide pXL545, qui ne permet que l'amplification de cobE, ne provoque pas d'augmentation d'activité SUMT. Ce résultat confirme que cobA est le gène de structure de la SUMT de Pseudomonas denitrificans. Ce résultat montre que l'on peut obtenir une amplification de l'activité SUMT chez Pseudomonas denitrificans par amplification du gène de structure de la SUMT de Pseudomonas denitrificans.

#### 7.1.2 - Expression de la protéine COBI

Un fragment provenant du fragment d'ADN de 8,7 kb, contenant le gène de structure de la SP<sub>2</sub>MT (cobI), est cloné sur un plasmide à large spectre d'hôte chez les bactéries gram-négatives, puis ce plasmide est ensuite introduit par conjugaison chez Pseudomonas denitrificans SC510 Rif<sup>r</sup>. L'activité S-adénosyl-L-méthionine:précorrine-2 méthyltransférase de la souche est ensuite mesurée par rapport à celle de la souche portant le vecteur.

Le fragment BamHI-BamHI-SstI-SstI de 1,9 kb contenant les gènes cobH et cobI est purifié à partir du fragment de 8,7 kb. Des "linkers" XbaI et EcoRI sont placés respectivement aux extrémités BamHI et SstI après que celles-ci aient été remplies avec l'ADN polymérase du bactériophage T4. Le fragment est ensuite inséré entre les sites XbaI et EcoRI du plasmide à large spectre d'hôte pXL59. Il

porte la résistance à la kanamycine. Le plasmide ainsi obtenu est nommé le pXL1148 (figure 24).

Par ailleurs un plasmide voisin a été construit : le fragment BamHI-BamHI-SstI de 1,5 kb contenant seulement le gène cobH en entier et la partie 5' du gène cobI a été purifié à partir du fragment de 8,7 kb. Des "linkers" XbaI et EcoRI ont été ajoutés aux sites BamHI et SstI respectivement après que ceux-ci aient été remplis ou digérés avec l'ADN polymérase du phage T4. Ce fragment a ensuite été inséré entre les sites EcoRI et XbaI du pXL59 pour donner le plasmide pXL1149. Les plasmides pXL1148 et pXL1149 ne diffèrent que par la présence sur pXL1148 du fragment SstI-SstI de 0,3 kb qui contient l'extrémité 3' du gène cobI. pXL1148 possède le gène de structure entier de cobI au contraire de pXL1149. Les deux plasmides contiennent le gène cobH.

Ces deux plasmides ont été introduits par conjugaison chez SC510 Rif<sup>r</sup>. Les souches SC510 Rif<sup>r</sup>, SC510 Rif<sup>r</sup> pXL59, SC510 Rif<sup>r</sup> pXL1148 et SC510 Rif<sup>r</sup> pXL1149 sont cultivées en milieu PS4. Après 4 jours de culture, les cellules sont récoltées et les activités SP<sub>2</sub>MT sont dosées comme cela est décrit dans l'exemple 6.1.3 a).

Le résultat de ces dosages est porté ci-dessous avec les activités de SP<sub>2</sub>MT définies comme à l'exemple 6.1.3 a).

Tableau : Activités SP<sub>2</sub>MT de diverses souches dérivées de Pseudomonas denitrificans

Souche	Activité SP <sub>2</sub> MT <sup>1</sup> en %
SC510 Rif <sup>r</sup>	< 5
SC510 Rif <sup>r</sup> pXL59	< 5
SC510 Rif <sup>r</sup> pXL1148	75
SC510 Rif <sup>r</sup> pXL1149	< 5

pour 500 µg d'extrait brut introduit dans le test.

L'activité est exprimée en % comme cela est défini à l'exemple 6.1.3 a).

Seul le plasmide pXL1148 apporte une augmentation sensible de l'activité SP<sub>2</sub>MT. Au contraire le plasmide pXL1149 ne donne pas de résultats différents de ceux observés avec les contrôles SC510 Rif<sup>r</sup> et SC510 Rif<sup>r</sup> pXL59. Le pXL1148 est le seul plasmide à contenir le gène cobI et il est le seul à amplifier l'activité SP<sub>2</sub>MT; ce résultat confirme que le gène de structure de la SP<sub>2</sub>MT de Pseudomonas denitrificans est le gène cobI. De plus, si les protéines totales de ces différentes souches sont séparées par électrophorèse en conditions dénaturantes (PAGE-SDS à 10 % d'acrylamide) on observe spécifiquement dans le cas du pXL1148 la présence d'une bande qui correspond à une protéine ayant un poids moléculaire de 25 000 environ (figure 25). Le poids moléculaire de cette protéine correspond à celui de la protéine COBI. Le plasmide pXL1148 permet d'obtenir chez Pseudomonas denitrificans la surproduction de la protéine COBI.

#### 7.1.3 - Expression de COBF

L'expression est obtenue en positionnant en amont du gène cobF le promoteur P<sub>trp</sub> de E.coli et le site de fixation des ribosomes du gène cII du bactériophage lambda. L'expression ainsi obtenue est beaucoup plus élevée que celle observée par simple amplification génique grâce au même plasmide multicopie.

Le fragment EcoRI-BamHI-BamHI de 2 kb environ du pXL1496 (exemple 7.2.1) est purifié (figure 26). Ce fragment contient le promoteur P<sub>trp</sub> de E.coli et le site de fixation des ribosomes du gène cII du bactériophage lambda en amont du gène cobF. En aval du gène cobF se trouve le terminateur de l'opéron rnbB de E.coli. Ce fragment est cloné aux sites EcoRI-BamHI du plasmide pKT230 pour donner le pXL1546 (figure 26). Le pKT230 est un plasmide du groupe d'incompatibilité Q qui se réplique chez presque toutes les bactéries gram-négatives (Bagdasarian et al., 1981); ce plasmide porte la résistance à la kanamycine. Le plasmide pXL1546 et pKT230 sont

introduits, par conjugaison, chez SC510 Rif<sup>r</sup>. Les souches SC510 Rif<sup>r</sup>, SC510 Rif<sup>r</sup> pKT230 et SC510 Rif<sup>r</sup> pXL1546 sont cultivées en milieu PS4 comme cela est décrit précédemment. Après quatre jours de culture, les protéines totales des différentes souches sont analysées en PAGE-SDS à 5 10 %. Comme cela est montré à la figure 27, on observe dans l'extrait de SC510 Rif<sup>r</sup> pXL1546 une protéine d'un poids moléculaire de 32 000 environ qui est surexprimée; cette protéine co-migre avec la protéine qui est surexprimée par E.coli B pXL1496 (exemple 7.2.1). De plus cette protéine est spécifiquement exprimée dans la souche SC510 Rif<sup>r</sup> 10 contenant le pXL1546 où elle représente au moins 20 % des protéines totales. Par contre on n'observe pas cette protéine dans les protéines totales des souches SC510 Rif<sup>r</sup> et SC510 Rif<sup>r</sup> pKT230. Cette protéine surexprimée est donc la protéine COBF.

#### 7.1.4 - Expression de COBH

15 Cet exemple décrit l'amplification d'un fragment d'ADN de Pseudomoas denitrificans contenant le gène cobH. La protéine qui est codée par ce gène est purifiée; il s'agit de la protéine COBH. Le plasmide pXL1149, décrit à l'exemple 7.1.2, ne contient sur l'insert d'ADN provenant du fragment de 8,7 kb que le gène cobH en entier. Chez 20 SC510 Rif<sup>r</sup>, ce plasmide, contrairement au vecteur, entraîne la surexpression d'une protéine de poids moléculaire 22 000 (figure 25).

#### 7.1.5 - Expression de COBV

Cet exemple décrit l'amplification de l'activité cobalamin (5'-phosphate) synthase par un plasmide ne portant que cobV (pXL699, voir figure 38) L'activité cobalamin (5'-phosphate) synthase est amplifiée chez SC877Rifr par le plasmide pXL699 d'un facteur 50 par rapport à la même souche avec le vecteur pXL435, pXL1303, pXL1324 ou pKT230. Ce plasmide ne contient sur son insert que cobV en entier plus les parties 5' terminales d'ORF18 et de cobU. Il est certain que dans 25 30 une telle souche (SC877Rifr pXL699) la protéine COBV est surexprimée; cette surexpression est d'un facteur 50 par rapport à l'expression de la souche SC877Rifr.

#### 7.1.6. Expression de la protéine CORA

Le fragment EcoRI-BamHI-BamHI de 1.5 kb du pXL1832 (voir exemple 7.2.4) contenant le promoteur P<sub>trp</sub> puis le RBS cII du bactériophage λ, le gène de structure de la SUMT de *M. ivanovii* et la 5 région terminatrice de l'opéron rrnB de *E. coli* est cloné aux sites EcoRI-BamHI du pKT230 (Bagdasarian et al., 1981). De cette manière, le plasmide pXL1841 est obtenu (voir figure 56). Ce plasmide est mobilisé chez *P. denitrificans* SC510 Rifr comme cela est décrit précédemment. Un transconjugant est étudié plus en détail. Cette souche est cultivée en 10 milieu PS4 et l'activité SUMT des extraits bactériens est dosée en même temps que celle de la souche contrôle SC510 Rifr pXL435 (Cameron et al., 1989). Les activités de ces souches sont présentées ci-dessous.

Souche	Activité spécifique SUMT en pmol/h/mg de protéines
15 SC510 Rifr pXL435	50-100
SC510 Rifr pXL1841	1700

Ce résultat montre clairement qu'il y a expression de l'activité SUMT de *M. ivanovii* chez *P. denitrificans* grâce au plasmide pXL1841 puisque l'activité SUMT de la souche SC510 Rifr pXL1841 est nettement supérieure 20 à celle de SC510 Rifr pXL435.

#### 7.2 - Expression chez *E.coli*

Cet exemple illustre comment une protéine COB de *Pseudomonas denitrificans* peut être surproduite chez *E.coli*.

##### 7.2.1 - Expression de COBF

La surproduction est obtenue en positionnant en amont du 25 gène cob le promoteur P<sub>trp</sub> de *E.coli* et le site de fixation des ribosomes du gène cII du bactériophage lambda. Le fragment EcoRI-XbaI de 2250 pb du fragment EcoRI de 8,7 kb (aux positions respectives 0 et

2250 sur la séquence présentée à la figure 8) a été cloné dans le phage M13mp19 (Norrander et al., 1983) entre les sites EcoRI et Sall. Le plasmide ainsi construit est nommé pXL1405. Un site NdeI a été introduit par mutagénèse dirigée de manière à ce que les trois dernières bases (ATG) de ce site de restriction constituent le site d'initiation de traduction du gène cobF. Cela revient à modifier les trois bases qui précèdent l'ATG du gène cobF GAA (le G est à la position 733 sur la séquence présentée sur la figure 8) en CAT. Le fragment NdeI-SphI-SphI (figure 26) contenant le gène cobF est ensuite purifié; ce fragment de 1,5 kb est ensuite cloné entre les sites NdeI-SphI du plasmide pXL694 (Denèfle et al., 1987). Le plasmide ainsi construit est nommé pXL1496 (figure 26). Sur le fragment EcoRI-NdeI de 120 pb (qui provient du pXL694) qui précède le gène cobF sont présents des signaux de régulation de l'expression génétique de E.coli. Ces signaux sont constitués par la région [-40+1] du promoteur Ptrp de E.coli, puis par 73 pb qui contiennent le site de fixation des ribosomes du gène cII du bactériophage  $\lambda$  (Denèfle et al., 1987). En aval du gène cobF se trouvent les terminateurs de l'opéron rrnB de E.coli (sur le fragment HindIII-BamHI). Le plasmide pXL1496 a été introduit par transformation dans la souche de E.coli (Monod et Wollman, 1947). L'expression du gène cobF a été étudiée comme cela est déjà décrit (Denèfle et al., 1987) dans des conditions où le promoteur Ptrp est soit réprimé (présence de tryptophane), soit non réprimé (absence de tryptophane). Le milieu où l'expression a été réalisée est du milieu minimum M9 (Miller, 1972) supplémenté par 0,4 % de glucose, 0,4 % de casaminoacides, 10 mM thiamine et 40  $\mu$ g/ml de tryptophane dans le cas où l'on veut réprimer le promoteur Ptrp. La souche E.coli B pXL1496 a été cultivée à 37°C dans le milieu décrit ci-dessus avec 100  $\mu$ g d'ampicilline. Comme il est montré sur la figure 28, l'absence de tryptophane entraîne l'expression d'une protéine d'un poids moléculaire de 32 000. En effet dans l'extrait de protéines totales de E.coli B pXL1496 analysé en PAGE-SDS (figure 28) on observe clairement une protéine d'un poids moléculaire de 32 000 qui représente entre 1 et 4 % des protéines totales. Cette protéine est présente en quantités nettement moins importantes dans l'extrait des protéines totales de

E.coli B pXL1496 cultivée dans les mêmes conditions, mais en présence de tryptophane. Le poids moléculaire de la protéine qui est exprimée dans ces conditions est proche du poids moléculaire de la protéine COBF déduit de la séquence en acides aminés de la protéine qui est 28 5 927 (figure 16). La protéine qui est ainsi exprimée chez E.coli est la protéine COBF.

#### 7.2.2 - Expression de COBT

La surproduction est obtenue en fusionnant à l'extrémité 5' du gène cob le promoteur lac et les trois premiers codons de lacZ 10 de E.coli.

Le site EcoRI situé à la position 2624 sur la séquence présentée sur la figure 32 du fragment de 4.8 kb contient le quatrième codon du gène cobT. Le fragment EcoRI-XbaI de 3,5 kb du pXL837 (voir figure 36) est cloné aux sites EcoRI et XbaI du pTZ18R ou pTZ19R (Pharmacia) pour 15 générer les pXL1874 ou pXL1875 respectivement; ces deux plasmides diffèrent par l'orientation du gène cobT tronqué vis à vis du promoteur de l'opéron lactose d'E. coli (Plac). Plac est en amont de cobT sur pXL1874 tandis que c'est le contraire sur le pXL1875. Le clonage aux sites EcoRI-XbaI du pTZ18R du fragment EcoRI-XbaI du 20 pXL837 permet de réaliser une fusion de protéine entre les 4 premiers acides aminés de la β-galactosidase d'E. coli et le gène cobT à partir de son 4ème codon. L'expression de ce gène lacZ' 'cobT est sous le contrôle des signaux d'expression de lacZ. Les plasmides pXL1874, 25 pXL1875, pTZ18R sont introduits par transformation dans la souche de E.coli BL21. L'expression du gène cobT est étudié comme cela est déjà décrit (Maniatis et al., 1989).

Comme le montre la figure 42B, une protéine d'un poids moléculaire de 72 000 n'est exprimée qu'avec le pXL1874 et représente, dans l'extrait de protéines totales de BL21, pXL1874 analysé en PAGE-SDS, 1 à 4 % des 30 protéines totales. Le poids moléculaire de la protéine qui est exprimée dans ces conditions est proche du poids moléculaire de la protéine COBT déduit de la séquence en acides aminés qui est de 70 335 sur la figure 40. Cette expérience montre clairement qu'à partir du site EcoRI situé sur le quatrième codon du gène cobT une phase ouverte

compatible avec celle trouvée pour le gène cobT peut être exprimée.

#### 7.2.3 - Expression d'une protéine COBS tronquée.

Un site BamHI est situé au niveau du 45ème codon du gène COBS. Le fragment BamHI-BamHI de 1.2 kb contenant la partie 3' du gène cobS et des séquences en aval de ce gène est excisé du pXL843 pour être cloné au site BamHI du plasmide pET-3b (Rosenberg et al., 1987) pour générer le pXL1937. Le fragment BamHI est orienté de manière à ce que la partie tronquée du gène cobS soit fusionnée, en phase, avec les 12 premiers codons de la protéine majeure de capsid du bactériophage T7 ou gène 10 (Rosenberg et al., -1987). Ce gène hybride est sous le contrôle du promoteur  $\phi$ 10 du bactériophage T7. Le plasmide pXL1937 ainsi que le pET-3b sont introduits par transformation dans la E. coli BL21 pLysS (W. Studier, communication personnelle). Après réisolement sur milieu sélectif, les 2 souches sont cultivées en milieu liquide L jusqu'à une D<sub>0</sub> 610 nm de 1; à ce stade là le milieu est ajusté à une concentration en IPTG (isopropyl  $\beta$ -galactoside) de 1 mM afin d'induire l'expression de la polymérase du bactériophage T7 (Rosenberg et al., 1987). La culture est ensuite incubée 3 h à 37°C, puis des lysats bactériens sont préparés. Les protéines totales des bactéries ainsi cultivées sont séparées par PAGE en conditions dénaturantes. Comme on le voit sur la figure 42A, il y a spécifiquement surexpression d'une protéine de 33 000 avec la culture BL21 pLysS pXL1937. Ce poids moléculaire est tout à fait compatible avec le poids moléculaire attendu pour la protéine fusion (33 kD). Cette expérience montre clairement qu'à partir du site BamHI situé au niveau du 45ème codon du gène cobS, une phase ouverte compatible avec celle trouvée pour le gène cobS peut être surexprimée.

#### 7.2.4: Expression de la protéine CORA

Les oligonucléotides suivants ont été synthétisés comme cela est 30 décrit précédemment :

oligonucléotide 1277

5' GGC CGA ATT CAT ATG GTA GTT TAT TTA 3'

-----1 2 3 4 5 (1 à 5, 5

EcoRI ----- premiers codons

NdeI de la SUMT de M. ivanovii)

5 oligonucléotide 1278

5' GGC CGA GCT CTA TTA CAT AAT T

=====

SstI

(= séquence figurant sur la figure 51, positions 926 à 915 sur le brin  
10 complémentaire du brin codant)

L'oligonucléotide 1277 possède les séquences de reconnaissance des enzymes de restriction EcoRI et NdeI. Les trois dernières bases du site NdeI (ATG), qui correspondent à un codon d'initiation de traduction, sont directement suivies par les codons 2 à 5 du gène de structure de la 15 SUMT de M. ivanovii tels qu'ils figurent dans la séquence présentée sur la figure 52. L'oligonucléotide 1278 contient la séquence de reconnaissance de SstI, suivie directement de la séquence TATTACATAATT qui correspond à une séquence présente sur le fragment de 955 pb contenant le gène corA présenté sur la figure 51; cette séquence se 20 trouve à la position 926 à 915 (voir figure 51) sur le brin complémentaire du brin codant pour la protéine CORA. Les deux oligonucléotides 1277 et 1278 contiennent donc dans leur partie 3' des séquences correspondant respectivement au brin codant du gène corA et au brin complémentaire en aval de ce gène. Ces deux oligonucléotides 25 peuvent être utilisés pour réaliser une expérience d'amplification enzymatique avec le plasmide pXL1809 comme matrice. Cette expérience permet d'obtenir un fragment de 910 pb contenant le gène corA de M. ivanovii possédant un site NdeI au niveau de l'ATG du gène corA et un site SstI à l'autre extrémité du fragment après la fin du gène corA. 30 L'amplification enzymatique est réalisée comme précédemment décrit pour l'amplification enzymatique faite sur l'ADN génomique de M. ivanovii, si ce n'est que la matrice est constituée par 10 ng d'ADN du plasmide pXL1809; les températures utilisées sont les mêmes, mais seulement 20

cycles d'amplification sont réalisés. Comme cela est précédemment décrit, les produits d'amplification sont digérés par NdeI et SstI avant d'être séparés par migration sur gel d'agarose. Comme attendu, un fragment d'une taille 910 pb est effectivement visualisé. Ce fragment 5 est purifié comme cela est déjà décrit. Ce fragment est cloné aux sites NdeI et SstI du pXL694 (Denèfle et al., 1987). Le plasmide résultant nommé pXL1832 est décrit sur la figure 56. Sur ce plasmide de la même manière que cela est décrit à l'exemple 7.2, le gène de structure de la SUMT de *M. ivanovii* est précédé par le site de fixation des ribosomes du 10 gène CII du bactériophage  $\lambda$ . En amont de ce RBS se trouve le promoteur P<sub>trp</sub>. Le plasmide pXL1832 est introduit dans *E. coli* B5548 qui est une souche de *E. coli* portant la mutation cysG44 (Cossart et Sanzey, 1982) par transformation. Les activités SUMT des souches *E. coli* B5548 pUC13 et *E. coli* B5548 pXL1832 sont dosées sur des extraits obtenus à partir 15 de cellules cultivées en milieu LB supplémenté en ampicilline. Le dosage de l'activité SUMT est réalisé comme cela a déjà été décrit (Blanche et al., 1989). Les résultats de ce dosage sont donnés ci-dessous.

Souche	Activité spécifique SUMT en pmol/h/mg de protéines
20 <i>E. coli</i> B5548 pUC13	5.9
<i>E. coli</i> B5548 pXL1832	310

Les résultats présentés sur le tableau ci-dessus montrent clairement qu'il y a expression d'une activité SUMT chez la souche *E. coli* B5548 lorsque celle-ci contient une plasmide pXL1832 qui exprime la SUMT de *M. ivanovii*. La SUMT de *M. ivanovii* peut donc être exprimée chez *E. coli*. 25

#### EXEMPLE 8 - Amplification de la production de cobalamines par les techniques d'ADN recombinant

##### 8.1 - Amplification chez *P.denitrificans*

Cet exemple illustre comment une amélioration de la 30 production de cobalamines est obtenue chez *Pseudomonas denitrificans*

SC510 Rif<sup>r</sup> par amplification de gènes cob de Pseudomonas denitrificans SC510.

8.1.1 Amélioration de la production de cobalamines chez Pseudomonas denitrificans par levée d'une étape limitante dans la biosynthèse des cobalamines

Cet exemple illustre comment la productivité en cobalamines de souches de Pseudomonas denitrificans peut être améliorée par amplification de gènes cob de Pseudomonas denitrificans. Cette amélioration résulte de la levée d'une étape limitante de la voie de biosynthèse.

Le plasmide pXL367 est décrit à l'exemple 4.2 (figure 13). Ce plasmide correspond au pRK290 (Ditta et al., 1981) dans lequel le fragment EcoRI de 8,7 kb a été inséré. Ce plasmide pXL367 procure une amélioration de la biosynthèse de cobalamines chez la souche SC510 Rif<sup>r</sup>. Les souches SC510 Rif<sup>r</sup>, SC510 Rif<sup>r</sup> pRK290 et SC510 Rif<sup>r</sup> pXL367 sont cultivées en erlenmeyer dans du milieu PS4 suivant les conditions décrites dans les protocoles expérimentaux. On observe une amélioration du titre de production due à la présence du plasmide pXL367. En effet la souche SC510 Rif<sup>r</sup> pXL367 produit 30 % de plus de cobalamines que les souches SC510 Rif<sup>r</sup> et SC510 Rif<sup>r</sup> pRK290. Cette amélioration n'est pas due à l'amplification de n'importe quels gènes de Pseudomonas denitrificans mais à l'amplification spécifique de gènes portés par le fragment EcoRI de 8,7 kb. En effet le plasmide pXL723 décrit sur la figure 11 ne donne aucune amélioration et le même titre de production est observé avec ce plasmide qu'avec les souches SC510 Rif<sup>r</sup> et SC510 Rif<sup>r</sup> pRK290.

8.1.2 Amélioration de la production de coenzyme B<sub>12</sub> chez Pseudomonas denitrificans par levée de deux étapes limitantes dans la biosynthèse des cobalamines

Cet exemple illustre comment la productivité en cobalamines de souches de Pseudomonas denitrificans peut être améliorée par amplification de gènes cob de Pseudomonas denitrificans. Cette amélioration résulte de la levée de deux étapes limitantes de la

voie de biosynthèse.

Le fragment ClaI-EcoRV de 2,4 kb issu du fragment de 5,4 kb (contenant les gènes cobA et cobE) est cocloné avec le fragment EcoRI de 8,7 kb sur le plasmide à large spectre d'hôte pXL203. Le 5 plasmide ainsi construit est appelé pXL525 (figure 29). Ce plasmide est introduit chez SC510 Rif<sup>r</sup> par conjugaison. La souche SC510 Rif<sup>r</sup> pXL525 produit 20 % de cobalamines de plus que SC510 Rif<sup>r</sup> pXL367. L'amplification des gènes cobA et cobE permet de lever une nouvelle étape limitante chez SC510 Rif<sup>r</sup> dans la biosynthèse des cobalamines. 10 La souche SC510 Rif<sup>r</sup> de Pseudomonas denitrificans est améliorée dans le présent exemple par la levée successive de deux étapes limitantes. Cet exemple montre que la levée de deux étapes limitantes dans la biosynthèse des cobalamines peut conduire à des améliorations supplémentaires de production.

15           8.2 - Amélioration de la productivité de cobalamines chez Agrobacterium tumefaciens

Cet exemple illustre l'amélioration de la production de cobalamines d'une souche productrice de cobalamines par l'amplification de gènes cob de Pseudomonas denitrificans SC510.

20           La souche utilisée est une souche de bactérie gram-négative ; il s'agit d'une souche d'Agrobacterium tumefaciens.

Les plasmides décrits dans les exemples 4.2 et 8.1, pXL367 et pXL525 ainsi que le vecteur pRK290 (Ditta et al., 1981) et le plasmide pXL368 (figure 29), sont introduits par transfert conjugalif 25 chez la souche d'Agrobacterium tumefaciens C58-C9 Rif<sup>r</sup> (Cameron et al., 1989). Les souches C58-C9 Rif<sup>r</sup>, C58-C9 Rif<sup>r</sup> pRK290, C58-C9 Rif<sup>r</sup> pXL367, C58-C9 Rif<sup>r</sup> pXL368 et C58-C9 Rif<sup>r</sup> pXL525 sont cultivées en milieu PS4 à 30°C comme cela est décrit précédemment. Les cobalamines produites sont dosées comme cela est décrit précédemment. Les titres 30 de productions sont portés sur le tableau ci-dessous.

Tableau : Titres de vitamine B<sub>12</sub> produite par différentes souches d'Agrobacterium tumefaciens recombinantes.

Souche	Vitamine B <sub>12</sub> en mg/l
C58-C9 Rif <sup>r</sup>	0.4
5 C58-C9 Rif <sup>r</sup> pRK290	0.4
C58-C9 Rif <sup>r</sup> pXL367	0.8
C58-C9 Rif <sup>r</sup> pXL368	0.8
C58-C9 Rif <sup>r</sup> pXL525	1.2

Comme il apparaît clairement sur le tableau ci-dessus, la production de cobalamines est améliorée chez la souche d'Agrobacterium tumefaciens utilisée. Deux plasmides différents améliorent la production de cobalamines de la souche d'Agrobacterium tumefaciens utilisée: pXL367 et pXL368. Ces plasmides contiennent respectivement le fragment EcoRI de 8,7 kb (gènes cobF à cobM) et le fragment ClaI-EcoRV de 2,4 kb (gène cobE et cobA). Séparément, ils améliorent la production de cobalamines d'Agrobacterium tumefaciens C58-C9 Rif<sup>r</sup> d'un facteur 2; ce résultat montre qu'il est possible d'améliorer la production de cobalamines d'une souche d'Agrobacterium tumefaciens en amplifiant des fragments portant des gènes cob de Pseudomonas denitrificans. Dans le cas présent on peut parler d'amélioration hétérologue, c'est-à-dire d'amélioration de la production de cobalamines d'une souche moyennant l'amplification de gènes cob d'une autre souche.

Les améliorations de production de cobalamines apportées par les différents fragments de Pseudomonas denitrificans contenant des gènes cob sont cumulables, c'est-à-dire qu'en mettant sur le même plasmide les deux fragments qui sont séparément clonés sur le pXL367 et le pXL368, on observe des améliorations additives. Le plasmide pXL525 apporte chez Agrobacterium tumefaciens C58-C9 Rif<sup>r</sup> une amélioration de production supérieure à celle apportée par chacun des fragments clonés séparément sur le même vecteur.

8.3 - Amélioration de la productivité de cobalamines chez  
Rhizobium meliloti

Cet exemple décrit l'amélioration de la production de cobalamines d'une autre souche productrice de cobalamines.

5 Le plasmide décrit à l'exemple 8.2, pXL368, ainsi que le vecteur pRK290 (Ditta et al., 1981) sont introduits par transfert conjugalif chez la souche Rhizobium meliloti 102F34 Rif<sup>r</sup> (Leong et al., 1982). Les transconjugants: 102F34 Rif<sup>r</sup>, 102F34 Rif<sup>r</sup> pRK290 et 102F34 Rif<sup>r</sup> pXL368 sont cultivés en milieu PS4 à 30°C comme cela est 10 décrit précédemment. Les cobalamines produites sont dosées comme cela est décrit précédemment. Les titres de productions sont portés sur le tableau ci-dessous.

Tableau : Titres des cobalamines produites par différentes souches de Rhizobium meliloti recombinantes

	Souche	Vitamine B <sub>12</sub> en mg/l
15	102F34 Rif <sup>r</sup>	0.4
20	102F34 Rif <sup>r</sup> pRK290	0.4
	102F34 Rif <sup>r</sup> pXL368	0.8

Comme il apparaît clairement sur le tableau ci-dessus, la production de cobalamines est améliorée chez la souche de Rhizobium meliloti utilisée. Le plasmide pXL368 améliore la production de cobalamines de la souche de Rhizobium meliloti utilisée. Ce plasmide 25 contient le fragment ClaI-EcoRV de 2,4 kb (gènes cobA et cobE); il améliore la production de cobalamines de Rhizobium meliloti 102F34 Rif<sup>r</sup> d'un facteur 2. Ce résultat montre qu'il est possible d'améliorer la production de cobalamines d'une souche de Rhizobium meliloti en amplifiant des fragments portant des gènes cob de Pseudomonas denitrificans. Dans le cas présent on peut parler d'amélioration hétérologue, c'est-à-dire d'amélioration de la production de cobalamines d'une souche moyennant l'amplification de gènes cob d'une 30

autre souche.

EXEMPLE 9 - Dosage des corrinoïdes et des descobaltocorrinoïdes dans les moûts et les cellules de souches productrices de corrinoïdes

5 Cet exemple illustre comment il est possible d'identifier et de doser les différents corrinoïdes et descobaltocorrinoïdes produits par différentes souches productrices de cobalamines. Ce dosage permet entre autre de doser le coenzyme B<sub>12</sub>.

10 Les moûts (ou les cellules seules) sont cyanurés comme cela a déjà été décrit (Renz, 1971). Après centrifugation, une aliquote du surnageant est passée à travers une colonne de DEAE-Sephadex qui est ensuite lavée extensivement avec du dihydrogenophosphate de potassium 1M. Les fractions recueillies sont regroupées et dessalées sur une cartouche Sep-Pak C-18 (Waters). Après évaporation et resuspension dans l'eau (100 µl à 1 ml suivant la 15 quantité de corrinoïdes présents), les corrinoïdes sont identifiés et dosés par CLHP sur une colonne de Nucléosil C-18 (Macherey-Nagel). La colonne est éluée à 1 ml/mn avec un gradient d'acétonitrile (de 0 % à 100 %) dans du tampon phosphate de potassium 0,1 M contenant 10 mM de KCN.

20 Les corrinoïdes sont visualisés en détection U.V. à 371 nm et/ou par la détection spécifique du <sup>57</sup>Co (si la culture a été effectuée en présence de <sup>57</sup>CoCl<sub>2</sub>) à l'aide d'un détecteur Berthold LB 505. Ils sont donc identifiés et par comparaison de leur temps de rétention avec des étalons. De même, les descobaltocorrinoïdes (acide hydrogénobyrrinique, acide hydrogénobyrrinique monoamide et acide hydrogénobyrrinique diamide) sont visualisés en détection U.V. à 330 nm. Par cette technique les intermédiaires suivants sont séparés: l'acide cobyrrinique, l'acide cobyrrinique monoamide, l'acide cobyrrinique diamide, l'acide cobyrrinique triamide, l'acide cobyrrinique tétraamide, l'acide cobyrrinique pentaamide, l'acide cobyrique, le cobinamide, le cobinamide phosphate, le GDP-cobinamide, la vitamine B<sub>12</sub>-phosphate et la vitamine B<sub>12</sub>. Les formes adénosylées de ces 25 produits sont aussi séparées et dosées par cette technique. Pour ce 30

faire, l'étape initiale de cyanuration est supprimée et la colonne CLHP est éluée avec du tampon dépourvu de KCN. Nous donnons à la figure 31 les temps de rétention de différents étalons séparés par ce système et identifié à la sortie de la colonne par absorbance U.V.

5 Un échantillon de la souche SC510 Rif<sup>R</sup> a été déposé le 30 janvier 1990 au Centraal Bureau voor Schimmelcultures à Baarn (Pays-Bas) où il a été enregistré sous la référence CBS 103.90.

REVENDICATIONS

1. Séquence d'ADN caractérisée en ce qu'elle code pour un polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides.
2. Séquence d'ADN selon la revendication 1 caractérisée en ce qu'elle est choisie parmi
  - 5 . les gènes cobA, cobB, cobC, cobD, cobE, cobF, cobG, cobH, cobI, cobJ, cobK, cobL, cobM, cobN, cobO, cobP, cobQ, cobS, cobT, cobU, cobV, cobW, cobX, corA, présentés aux figures 15, 16, 40, 41, 47 et 52,
  - 10 . les homologues de ces séquences résultant de la dégénérescence du code génétique,
  - . les séquences d'origine naturelle, synthétique ou recombinante, qui hybrident et/ou qui présentent des homologies significatives avec ces séquences d'ADN ou avec des fragments de celles-ci, et qui codent pour
    - 15 des polypeptides impliqués dans la biosynthèse des cobalamines et/ou des cobamides.
3. Gène contenant une séquence d'ADN selon l'une des revendications précédentes.
4. ADN recombinant caractérisé en ce qu'il contient au moins une séquence d'ADN selon l'une des revendications 1 à 3.
- 20 5. ADN recombinant selon la revendication 4 caractérisé en ce que lesdites séquences d'ADN sont placées sous le contrôle de signaux d'expression.
- 25 6. ADN recombinant selon la revendication 5 caractérisé en ce que les signaux d'expression peuvent être homologues ou hétérologues de la séquence d'ADN.
7. ADN recombinant selon l'une des revendications 4 à 6 caractérisé en

ce qu'il fait partie d'un plasmide d'expression.

8. Plasmide caractérisé en ce qu'il contient au moins une séquence d'ADN codant pour un polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides, et des séquences permettant leur expression.

9. Plasmide selon la revendication 8 caractérisé en ce qu'il contient

- . un ADN recombinant selon l'une des revendications 4 à 7,
- . un système de réPLICATION,
- . au moins un marqueur de sélection.

10 10. Plasmide selon la revendication 8 caractérisé en ce qu'il est choisi parmi

- . le plasmide pXL1500 contenant les gènes cobA, cobB, cobC et cobE,
- . les plasmides pXL723 et pXL302 contenant les gènes cobC et cobD,
- . le plasmide pXL1397 contenant les gènes cobB et cobC,
- 15 . les plasmides pXL368 et pXL557 contenant les gènes cobA et cobE,
- . les plasmides pXL545 et pXL545 a contenant le gène cobE,
- . les plasmides pXL367 et pXL253 contenant les gènes cobF, cobG, cobH, cobI, cobJ, cobK, cobL, cobM,
- . le plasmide pXL1148 contenant les gènes cobH et cobI,
- 20 . le plasmide pXL1149 contenant le gène cobH,
- . les plasmides pXL1496 et pXL1546 contenant le gène cobF,
- . le plasmide pXL525 contenant les gènes cobA, cobE et cobF à cobM, le plasmide pXL843 contenant les gènes cobX, cobS, et cobT, le plasmide pXL699 contenant le gène cobV, le plasmide pXL1324 contenant le gène 25 cobU, les plasmides pXL618 et pXL623 contenant respectivement les gènes cobQ et cobP, le plasmide pXL593 contenant le gène cobP et cobW et le plasmide pXL1909 contenant les gènes cobP, cobW, cobN, et cobO.

11. Cellule dans laquelle a été introduite une séquence d'ADN selon l'une des revendications 1 à 7 ou un plasmide selon l'une des 30 revendications 8 à 10.

12. Polypeptide caractérisé en ce qu'il est impliqué dans la

biosynthèse des cobalamines et/ou des cobamides.

13. Polypeptide selon la revendication 12 caractérisé en ce qu'il est codé par une séquence d'ADN selon l'une des revendications 1 à 7.

14. Polypeptide selon la revendication 12 caractérisé en ce qu'il intervient dans la transformation du précorrine-3 en 5'-déoxy 5'-adénosyl(Ado) acide cobyrinique a,c-diamide.

15. Polypeptide selon la revendication 14 caractérisé en ce qu'il contient tout ou partie des séquences peptidiques COBB, COBF, COBG, COBH, COBJ, COBK, COBL, COBM, COBN, COBO, COBS et COBT présentées aux figures 15, 16, 40 et 41.

16. Polypeptide selon la revendication 14 caractérisé en ce qu'il catalyse le transfert d'un groupement méthyl aux positions C1, C5, C11, C15, ou C17 intervenant entre le précorrine-3 et l'acide cobyrinique a,c-diamide.

17. Polypeptide selon la revendication 16 caractérisé en ce qu'il contient tout ou partie des séquences peptidiques COBF, COBJ, COBL et COBM présentées à la figure 16.

18. Polypeptide selon la revendication 12 caractérisé en ce qu'il intervient dans la transformation de l'acide cobyrique en cobinamide.

19. Polypeptide selon la revendication 18 caractérisé en ce qu'il contient tout ou partie des séquences peptidiques COBC et COBD présentées à la figure 15.

20. Polypeptide selon la revendication 12 caractérisé en ce qu'il possède une activité S-adénosyl-L-méthionine:précorrine-2 méthyl transférase (SP2MT).

21. Polypeptide selon la revendication 20 caractérisé en ce qu'il

contient tout ou partie de la séquence peptidique COBI présentée à la figure 16.

22. Polypeptide selon la revendication 12 caractérisé en ce qu'il possède une activité acide cobyrinique et/ou hydrogénobyrinique  
5 a,c-diamide synthase.

23. Polypeptide selon la revendication 22 caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBB présentée à la figure 15.

24. Polypeptide selon la revendication 12 caractérisé en ce qu'il possède une activité precorrin-8x mutase.  
10

25. Polypeptide selon la revendication 24 caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBH présentée à la figure 16.

26. Polypeptide selon la revendication 12 caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBE présentée à la  
15 figure 15.

27. Polypeptide selon la revendication 12 caractérisé en ce qu'il possède une activité nicotinate-nucléotide: dimethylbenzimidazole phosphoribosyltransferase.

20 28. Polypeptide selon la revendication 27 caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBU présentée à la figure 41.

29. Polypeptide selon la revendication 12 caractérisé en ce qu'il possède une activité cobalamine-(5'-phosphate) synthase.

25 30. Polypeptide selon la revendication 29 caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBV présentée à la

figure 41.

31. Polypeptide selon la revendication 12 caractérisé en ce qu'il possède une activité cobyric acid synthase.
32. Polypeptide selon la revendication 31 caractérisé en ce qu'il 5 contient tout ou partie de la séquence peptidique COBQ présentée à la figure 47.
33. Polypeptide selon la revendication 12 caractérisé en ce qu'il possède une activité cob(I)alamin adénosyltransferase.
34. Polypeptide selon la revendication 33 caractérisé en ce qu'il 10 contient tout ou partie de la séquence peptidique COBO présenté à la figure 47.
35. Polypeptide selon la revendication 12 caractérisé en ce qu'il possède une activité précorrin-6x reductase.
36. Polypeptide selon la revendication 35 caractérisé en ce qu'il 15 contient tout ou partie de la séquence peptidique COBK présentée à la figure 16.
37. Polypeptide selon la revendication 12 caractérisé en ce qu'il intervient dans la transformation du cobinamide en GDP-cobinamide.
38. Polypeptide selon la revendication 37 caractérisé en ce qu'il 20 possède une activité cobinamide kinase et cobinamide phosphate guanylyltransferase.
39. Polypeptide selon la revendication 38 caractérisé en ce qu'il contient tout ou partie de la séquence peptidique COBP présentée à la figure 47
- 25 40. Polypeptide selon la revendication 12 caractérisé en ce qu'il

contient tout ou partie des séquences peptidiques COBS, COBT et COBX présentées figure 40.

41. Polypeptide selon la revendication 13 caractérisé en ce qu'il est choisi parmi les protéines COBA, CORA, COBB, COBC, COBD, COBE, COBF, 5 COBG, COBH, COBI, COBJ, COBK, COBL, COBM, COBN, COBO, COBP, COBQ, COBS, COBT, COBU, COBV, COBW et COBX présentées aux figures 15, 16, 40, 41, 47 et 52.

42. Procédé de production des polypeptides selon les revendications 12 à 41 caractérisé en ce que

- 10 . on introduit dans une cellule hôte une séquence d'ADN selon les revendications 1 à 7, ou un plasmide selon les revendications 8 à 10 contenant une telle séquence,
- . on cultive cette cellule recombinante dans des conditions d'expression de ladite séquence, et
- 15 . on récupère les polypeptides produits.

43. Procédé selon la revendication 42 caractérisé en ce que la cellule hôte peut être choisie parmi les procaryotes, les eucaryotes, les cellules animales ou végétales.

44. Procédé selon la revendication 43 caractérisé en ce que la cellule hôte est une archaebactérie ou une eubactérie.

45. Procédé selon la revendication 44 caractérisé en ce que la cellule hôte est E.coli, Pseudomonas denitrificans, Rhizobium meliloti, Agrobacterium tumefaciens, ou Salmonella typhimurium.

46. Procédé permettant d'augmenter la production de cobalamines et/ou 25 de cobamides ou de leurs précurseurs caractérisé en ce que . on introduit dans un microorganisme producteur de ces composés, ou potentiellement producteur de ces composés, une ou plusieurs séquences d'ADN codant pour un polypeptide impliqué dans la biosynthèse des cobalamines et/ou des cobamides,

- . on cultive le microorganisme ainsi obtenu dans des conditions de synthèse des cobalamines et/ou des cobamides et d'expression de ladite séquence,
- 5 . on récupère les cobalamines et/ou les cobamides ou leurs précurseurs produits.

47. Procédé selon la revendication 46 caractérisé en ce que l'on introduit dans le microorganisme une ou plusieurs séquences d'ADN selon l'une des revendications 4 à 7, ou un plasmide selon les revendications 8 à 10 contenant de telles séquences.

10 48. Procédé selon la revendication 47 caractérisé en ce que la séquence d'ADN introduite dans le microorganisme code pour un polypeptide catalysant une étape limitante de la biosynthèse des cobalamines et/ou des cobamides.

15 49. Procédé selon la revendication 47 caractérisé en ce que les séquences d'ADN introduites dans le microorganisme codent pour des polypeptides catalysant des étapes limitantes de la biosynthèse des cobalamines et/ou des cobamides.

20 50. Procédé selon l'une des revendications 46 à 49 caractérisé en ce que le microorganisme est choisi parmi P. denitrificans, R. melitoti, A. tumefaciens.

51. Procédé selon la revendication 50 caractérisé en ce que le microorganisme est P. denitrificans.

52. Procédé selon la revendication 51 caractérisé en ce que le microorganisme est P. denitrificans SC510 RifR.

25 53. Procédé selon l'une des revendications 46 à 52 caractérisé en ce que l'on introduit dans le microorganisme un plasmide selon la revendication 10.

54. Procédé selon les revendications 46 à 53 caractérisé en ce que l'on introduit dans la souche P. denitrificans SC510 RifR le plasmide pXL525.

5 55. Procédé selon les revendications 46 à 54 caractérisé en ce que les cobalamines et/ou les cobamides produits sont récupérés par  
. solubilisation,  
. conversion en forme cyano, et  
. purification.

10 56. Procédé selon les revendications 46 à 55 caractérisé en ce que la cobalamine est le coenzyme B12.

57. Procédé selon les revendications 46 à 55 caractérisé en ce que le précurseur est choisi parmi les descobaltocorrinoïdes et les corrinoïdes.

## Références bibliographiques.

5 Ausubel F. M., Brent R., Kinston R. E., Moore D. D., Smith J. A., Seidman J. G. and K. Struhl. 1987. Current protocols in molecular biology 1987-1988. John Wiley and Sons, New York.

10 Bagdasarian, M., R. Lurz, B. Rückert, F. C. Franklin, M. M. Bagdasarian, J. Frey, and K. Timmis. 1981. Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host vector system for gene cloning in *Pseudomonas*. *Gene* 16:237-247.

15 Barrère G, Geneste B., Sabatier A., 1981. Fabrication de la vitamine B12: l'amélioration d'un procédé. *Pour la Science*, 49, 56-64.

20 Battersby A. R., Fookes C. J. R., Matcham G. W. J., MacDonald E., 1980. Biosynthesis of the pigments of life: formation of the macrocycle. *Nature*, 285, 17-21.

25 Battersby, A. R., and E. MacDonald. 1982. Biosynthesis of the corrin macrocycle. p. 107-144. In D. Dolphin (ed.), B12, vol. 1. John Wiley & Sons, Inc., New-York.

30 Beck., W.S. 1982. Biological and medical aspects of vitamin B12. p 1-30. In D. Dolphin (ed.), B12, vol. 1. John Wiley & Sons, Inc., New-York.

35 Ben Bassat A., and K. Bauer. 1987. Amino-terminal processing of proteins. *Nature*, 326:315.

40 Blanche F., L. Debussche, D. Thibaut, J. Crouzet and B. Cameron. 1989. Purification and Characterisation of S-Adenosyl-L-Methionine:Uroporphyrinogen III methyltransferase from *Pseudomonas denitrificans*. *J. Bacteriol.*, 171:4222-4231.

45 Brey R. N., Banner C. D. B., Wölf J. B., 1986. Cloning of Multiple Genes Involved with Cobalamin (Vitamin B12) Biosynthesis in *Bacillus megaterium*. *J. Bacteriol.*, 167, 623-630.

50 Cameron B.;, K. Briggs, S; Pridmore, G. Brefort and J. Crouzet, 1989. Cloning and analysis of genes involved in coenzyme B12 biosynthesis in *Pseudomonas denitrificans* . *J. Bacteriol.*, 171, 547-557.

55 Casadaban, M. J., A. Martinez-Arias, S. T. Shapira and J. Chou. 1983.  $\beta$ -galactosidase gene fusion for analysing gene expression in *Escherichia coli* and Yeast. *Methods Enzymol.* 100, 293-308.

60 De Bruijn F. J. and J. R. Lupski. 1984. The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids-a review. *Gene*, 27, 131-149.

65 De Graff, J., J. H. Crosa, F. Heffron, and S. Falkow. 1978. Replication of the nonconjugative plasmid RSF1010 in *Escherichia coli* K-12. *J. Bacteriol.* 146, 117-122.

70 Denèfle P., S. Kovarik, J.-D. Guiton, T. Cartwright and J.-F. Mayaux. 1987. Chemical synthesis fo a gene coding for human angiogenin, its expression in *Escherichia coli* and conversion of the product into its active form. *Gene*, 56, 61-70.

- 141 -

5 Ditta G., Schmidhauser T., Yakobson E., Lu P., Liang X.-W., Finlay D. R., Guiney D. and D. R. Helinski. 1985. Plasmids related to the broad host range vector pRK290, useful for gene cloning and for monitoring gene expression. *Plasmid*, 13, 149-154.

10 Ditta, G., S. Stanfield, D. Corbin, and D. R. Helinski. 1980. Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene library of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* 77, 7347-7351.

15 Escalante-Semerena J. C. and J. R. Roth. 1987. Regulation of the cobalamin biosynthetic operons in *Salmonella typhimurium*. *J. Bacteriol.*, 169, 225-2258.

20 Florent, J. 1986. Vitamins. p115-158. In H.-J. Rehm and G. Reed (ed.), *Biotechnology*, vol.4, VCH Verlagsgesellschaft mbH, Weinheim.

25 Friedmann H. C. and L. M. Cagen. 1970. Microbial biosynthesis of B12-like compounds. *Ann. Rev. Microbiol.*, 24, 159-208.

30 Friedmann H. C., 1968. Vitamin B12 biosynthesis. *J. Biol. Chem.*, 243, 2065-2075.

35 Friedmann H. C., 1975. Biosynthesis of corrinoids. In Babior B. M., Cobalamin, 75-110, John Wiley and Sons, New York.

40 Henikoff S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene*, 28, 351-359.

45 Hirel Ph-H, J.-M. Schmitter, P. Dessen and S. Blanquet. 1989. Extent of N-terminal methionine excision within *E. coli* proteins is governed by the side chain of the penultimate aminoacids. *Proc. Natl. Acad. USA*, sous presse.

50 Hopp T. P. and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acids sequences. *Proc. Natl. Acad. Sci. USA*, 78-3824-3828.

55 Huennekens F. M., Vitols K. S., Fujii K., Jacobsen D. W., 1982. Biosynthesis of cobalamin coenzyme. In Dolphin D., B12, vol. 1, 145-167, John Willey & Sons, New York.

60 Irion R., Ljungdahl L. G., 1965. Isolation of factor IIIm coenzyme and cobyrinic acid coenzyme plus other B12 factors from *Clostridium thermoaceticum*. *Biochemistry*, 4, 2780-2790.

50 Jeter R. M., Olivera B. M., Roth J. R., 1984. *Salmonella typhimurium* synthetises cobalamin (vitamin B12) de novo under anaerobic growth conditions. *J. Bacteriol.*, 159, 206-213.

55 Jeter, R. M. and J. R. Roth. 1987. Cobalamin (Vitamin B12) Biosynthetic Genes of *Salmonella typhimurium*. *J. Bacteriol.* 169, 3189-3198.

60 Jorgensen R. A., Rothstein S. J., Reznikoff W. R., 1979. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. *Molec. Gen. Genet.*, 177, 65-62.

- 142 -

Kanangara C. G., S. P. Gough, P. Bruyant, J. K. Hoober, A. Kahn and D. von Wettstein. 1988. tRNAGlu as a cofactor in d-aminolevulinate biosynthesis: steps that regulate chlorophyll synthesis. *Trends in Biochem. Sci.*, 139-143.

5 Kanehisa M. 1984. Use of statistical criteria for screening potential homologies in nucleic acids sequences. *Nucleic Acids Res.*, 12:203-215.

10 Kiény M. P., R. Lathe and J. P. Lecocq. 1983. New versatile cloning vectors based on bacteriophage M13. *Gene*, 26, 91-99.

Krzycki J. and J. G. Zeikus. Quantification of corrinoids in methanogenic bacteria. 1980. *Curr. Microbiol.*, 3, 243-245.

15 L. Skatrud, A. J. Tietz, T. D. Ingolia, C. A. Cantwell, D. L. Fisher, J. L. Chapman and S. W. Queener. 1989. Use of recombinant DNA to improve production of cephalosporin C by Cephalosporium acremonium. *Bio/Technology*, 1989, 7, 477-485.

20 Laemmli U. K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680-685.

Leong S. A., Ditta G. S., Helinski D. R., 1982. Heme Biosynthesis in Rhizobium. Identification of a cloned gene coding for d-aminolevulinic acid synthetase from Rhizobium meliloti. *J. Biol. Chem.*, 257, 8724-8730.

25 Macdonald H. and J. Cole. Molecular cloning and functional analysis of the cysG and nirB genes of E. coli K12, Two closely-linked genes required for NADH-dependant reductase activity. submitted to publication.

30 Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

35 Mazumder T. K., N. Nishio, M. Hayashi and S. Nagai. 1987. Production of corrinoids including vitamin by Methanosarcina barkeri. 1986. *Biotechnol. Letters*, 12, 843:848.

40 Mazumder T. K., N. Nishio, S. Fukuzaki and S. Nagai. 1987. Production of Extracellular vitamin B12 compounds from methanol by Methanosarcina barkeri. *Appl. Microbiol. Biotechnol.*, 26, 511-516.

45 Miller, J. H. 1972. Experiment in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New-York.

50 Monod J. and E. Wollman. 1947. Inhibition de la croissance et de l'adaptation enzymatique chez les bactéries infectées par le bactériophage. *Ann. Inst Pasteur*, 73, 937-956.

Murphy M. J., Siegel L. M., Kamen H., Rosenthal D., 1973. Identification of a new class of heme prosthetic group: an iron-tetrahydroporphyrin (isobacteriochlorin type) with eighth carboxilic acid groups. *J. Biol. Chem.*, 248, 2801-2814.

55 Murphy M. J., Siegel L. M., 1973. The basis for a new type of porphyrin-related prosthetic group common to both assimilatory and dissimilatory sulfite reductases. *J. Biol. Chem.*, 248, 6911-6919.

60

Nexo E., Olesen H., 1982. Intrinsic factor, transcobalamin and haptocorrin. In Dolphin D., B12, 57-85, John Wiley & Sons, New York.

5 Normark S., S. Bergström, T. Edlund, T. Grundström, B. Jaurin, F. Lindberg and O. Olsson. 1983. Overlapping genes. Ann. Rev. Genet., 17, 499-525.

10 Norrander J., T. Kempe and J. Messing. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26, 101-106.

15 Noyes R., 1970. Vitamin B12 manufacture, 145-182, Noyes developpement S.A., Park Ridge, N. J., USA.

20 Prentki P. and H. M. Krisch. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene, 29, 303-313.

Renz P. 1970. Some intermediates in the biosynthesis of vitamin B12. Methods in Enzymol., 18, 82-92.

25 Rigby P. W. J., Dieckmann M., Rhodes C., Berg P., 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol., 113, 237.

Roof D. M. and J. R. Roth. 1988. Ethanolamine utilization in Salmonella typhimurium. J. Bacteriol., 170, 3855-3863.

30 Sanger F., S. Nicklen and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci., 74, 5463-5468.

Saunders G., Tuite M. F., Holt G., 1986. Fungal cloning vectors. Trends Biotechnol., 4, 93-98.

35 Scherer P., Höllriegel V., Krug C., Bokel M., Renz P., 1984. On the biosynthesis of 5-hydroxybenzimidazolylcobamide (vitamin B12-factor III) in Methanosaarcina barkeri. Arch. Microbiol., 138, 354-359.

40 Schneider Z., Friedmann H., 1972. Studies on enzymatic dephosphorylation of vitamin B12 5'-phosphate. Arch. Biochem. Biophys., 152, 488-495.

45 Scott A. I., N. E. Mackenzie, P. J. Santander, P. E. Fagerness, G. Muller, E. Schneider, R. Seldmeier, and G. Worner. 1984. Biosynthesis of vitamin B12-Timing of the methylation steps between uro'gen III and cobyrinic acid. Bioorg. Chem. 12:356-352.

50 Southern E., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., 98, 503-517.

55 Stachel S. E., G. An, C. Flores and E. W. Nester. 1985. A Tn3lacZ transposon for the random generation of  $\beta$ -galactosidase gene fusions: application to the analysis of gene expression in Agrobacterium. Embo J., 4, 891-898.

60 Staden R. and A. D. McLachlan. 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acid Res., 10, 141-156.

- 144 -

Stupperich E., I. Steiner and H. J. Eisinger. 1987. Substitution of Coa-(5-Hydroxybenzimidazolyl)Cobamide (Factor III) by vitamin B12 in Methanobacterium thermoautotrophicum. *J. Bacteriol.*, 169:3076-3081.

5 Taylor J. W., J. Ott and F. Eckstein. 1985. The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. *Nucl. Acid Res.*, 13, 8764-8765.

10 Viera J., Messing J., 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*, 19, 259-268.

Wein-Hsiung L., L. Chi-Cheng and W. Chung-I. 1985. Evolution of DNA sequences. p 1-94. In R. J. MacIntyre (ed.), *Molecular Evolutionary genetics*. Plenum Press, New York and London.

15 Latta, M., M. Philit, I. Maury, F. Soubrier, P. Denèfle and J.-F. Mayaux. 1990. Tryptophan promoter derivatives on multicopy plasmids : a comparative analysis of the expression potentials en *Escherichia coli*. *DNA Cell Biol.*, 9, 129-137.

Mayaux, J.-F., E. Cerbalaud, F. Soubrier, D. Faucher and D. Pétré. 1990. Purification, cloning and primary structure of an enantio-selective amidase from Brevibacterium sp. R312. Structural evidence for a genetic coupling with nitrile-hydrolase. 1990. *J. Bacteriol.*, 172, 6764-6773.

Belyaev, S. S., R. Wolkin, W. R. Kenealy, M. J. De Niro, M. J. Epstein and J. G. Zeikus. 1983. Methanogenic bacteria from Bondyuzhskoe oil field : general characterization and analysis of stable-carbon isotopic fractionation. *Appl. Environ. Microbiol.*, 45, 691-697.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239, 487-491.

Souillard, N., M. Magot, O. Possot and L. Sibold. 1988. Nucleotide sequence of regions homologous to NifH (nitrogenase Fe protein) from the nitrogen fixing archaebacteria Methanococcus thermolithotrophicum and Methanobacterium ivanovi : evolutionary implications. *J. Mol. Evol.*, 2, 65-76.

Chen, E. L. and P. H. Seeburg. 1985. Supercoil sequencing : a fast and simple method for sequencing plasmid DNA. *DNA*, 4, 165-170.

Saiki R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239, 487-491.

Grunstein M., Hogness D., 1975. Colony hybridisation: a method for the isolation of cloned DNAs that contains a specific gene. *Proc. Natl. Acad. Sci. USA*, 72, 3961-3971.

Cossart, P. and B. Gicquel-Sanzey. 1982. Cloning and sequence of the *cpr* gene of *Escherichia coli* K 12. *Nucleic Acid Res.*, 10, 1363-1378.

Viera, J. and J. Messing. 1987. Production of single stranded plasmid DNA. *Meth. Enzymol.*, 153, 3-11.

- 145 -

Barbieri P., Boretti A., Di Marco A., Migliacci A., and Spalla C. 1962. Further observations on the biosynthesis of vitamin B12 in Nocardia rugosa. *Biochim. Biophys. Acta.*, 57, 599-600.

Renz P. 1968. Reaktionfolge der enzymatischen synthese von vitamin B12 aus cobinamid bei Propionibacterium shermanii. *Z. Physiol. Chem.*, 349, 979-981.

Ronzio R. A., and Barker H. A. 1967. Enzimic synthesis of guanosine diphosphate cobinamide by extracts of propionic acid bacteria. *Biochemistry*, 6, 2344-2354.

Thibaut D., Debussche L., and Blanche F. 1990. Biosynthesis of vitamin B12: Isolation of precorrin-6x, a metal-free precursor of the corrin macrocycle retaining five S-adenosylmethionine-derived peripheral methyl groups. *Proc. Natl. Acad. Sci.*, 87, 8795-8799.

Ohta H., and Beck W. S. 1976. Studies of the ribosome-associated vitamin B12s adenosylating enzyme of *Lactobacillus leichmannii*. *Arch. Biochem. Biophys.*, 174, 713-725.

Brady R. O., Castanera E. G., and Barker H. A. 1962. The enzymatic synthesis of cobamide coenzymes. *J. Biol. Chem.*, 237, 2325-2332.

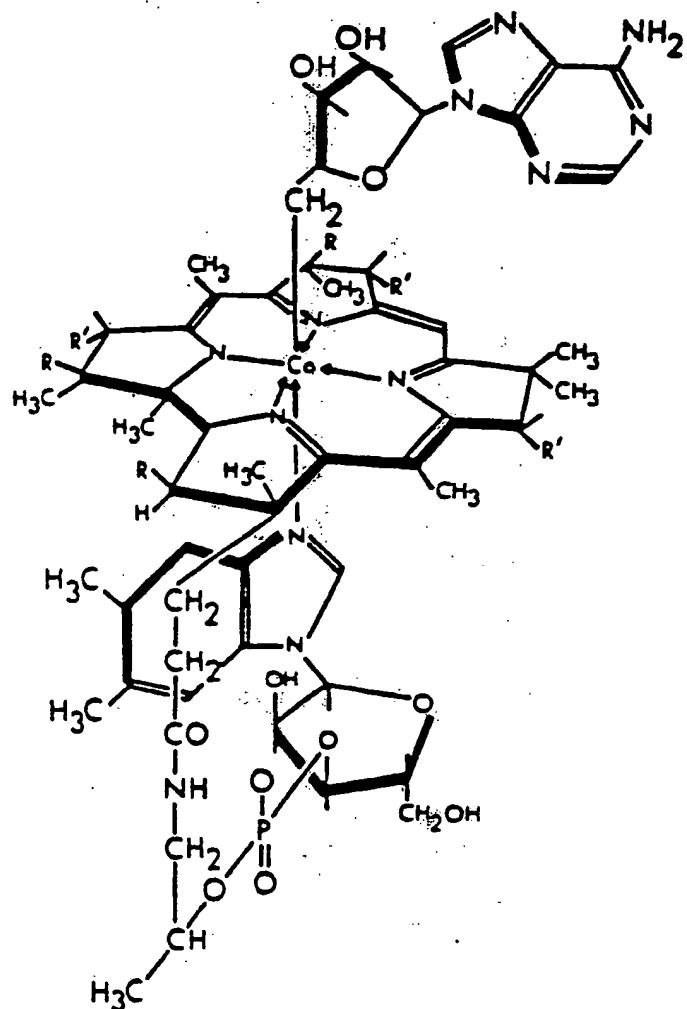
Fenton W. A., and Rosenberg L. E. 1978. Mitochondrial metabolism of hydroxocobalamin: synthesis of adenosylcobalamin by intact rat liver mitochondria. *Arch. Biochem. Biophys.*, 189, 441-447.

Vitols E., Walker G. A., and Huennekens F. M. 1966. Enzymatic conversion of vitamin B12s to a cobamide coenzyme,  $\alpha$ -(5,6-dimethyl-benzimidazolyl)deoxyadenosylcobamide (Adenosyl-B12). *J. Biol. Chem.*, 241, 1455-1461.

Gimsing P., and Beck W. S. 1986. Determination of cobalamins in biological material. *Methods Enzymol.*, 123, 3-14.

Jacobsen W. J., Green R., and Brown K. L. 1986. Analysis of cobalamin coenzymes and others corrinoids by high-performance liquid chromatography. *Methods Enzymol.*, 123, 14-22.

1/151



Adenosylcobalamin R = CH<sub>2</sub>CONH<sub>2</sub>, R' = CH<sub>2</sub>CH<sub>2</sub>CONH<sub>2</sub>.

FIGURE 1

.. 2 / 151

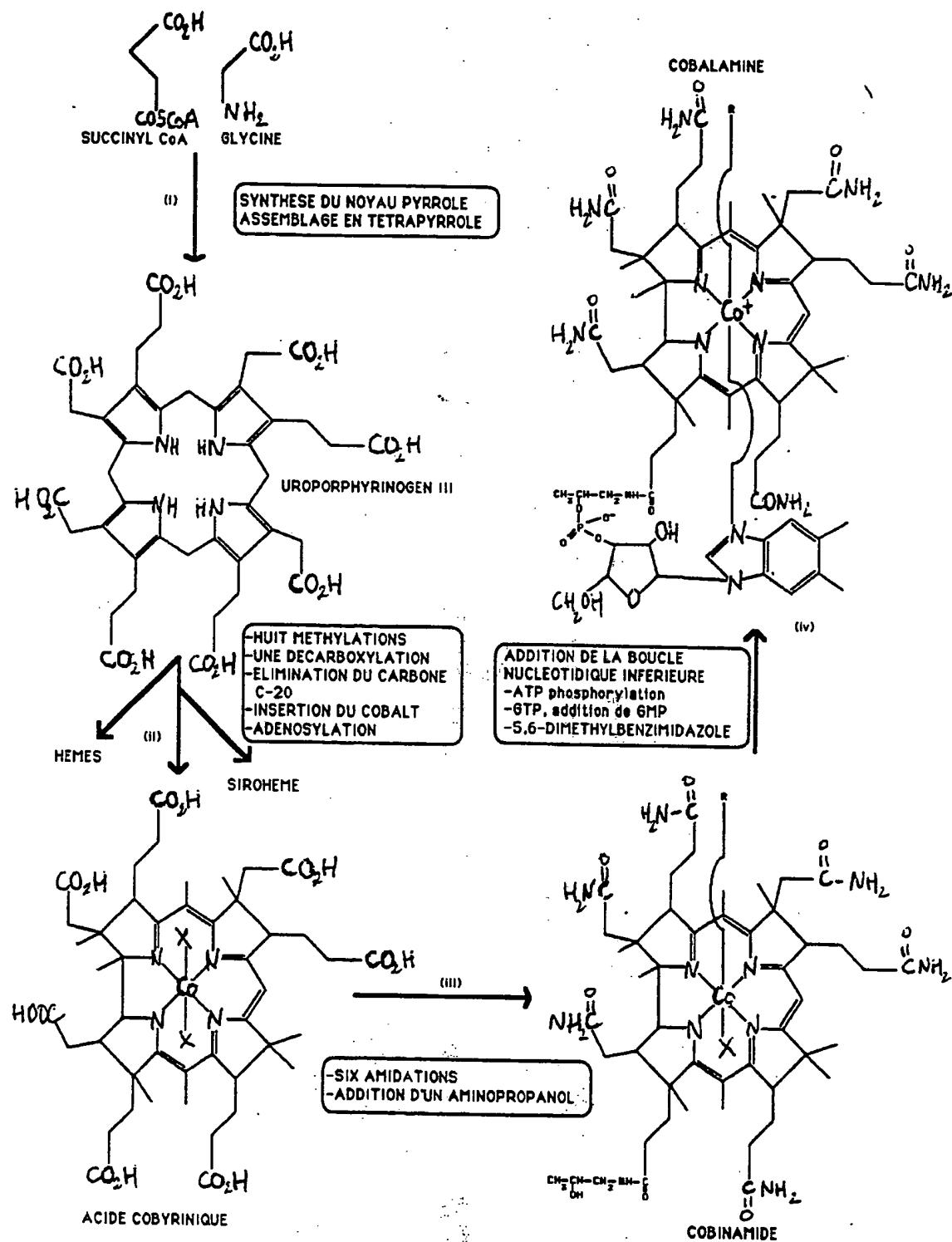
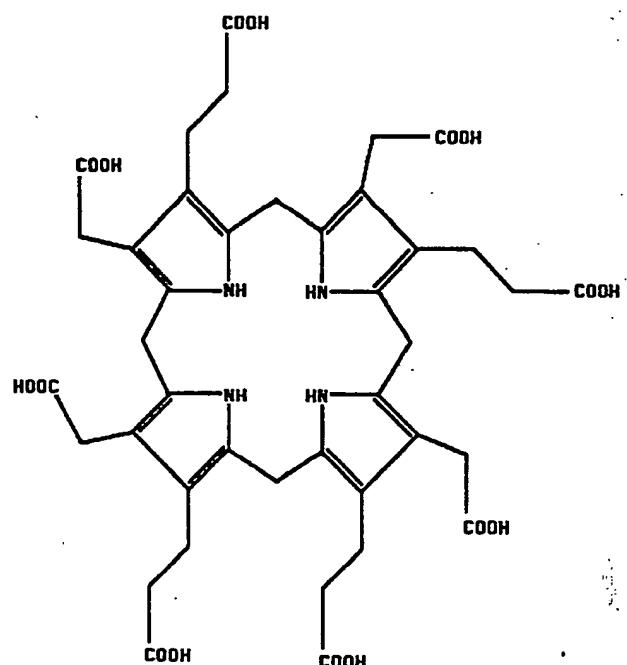
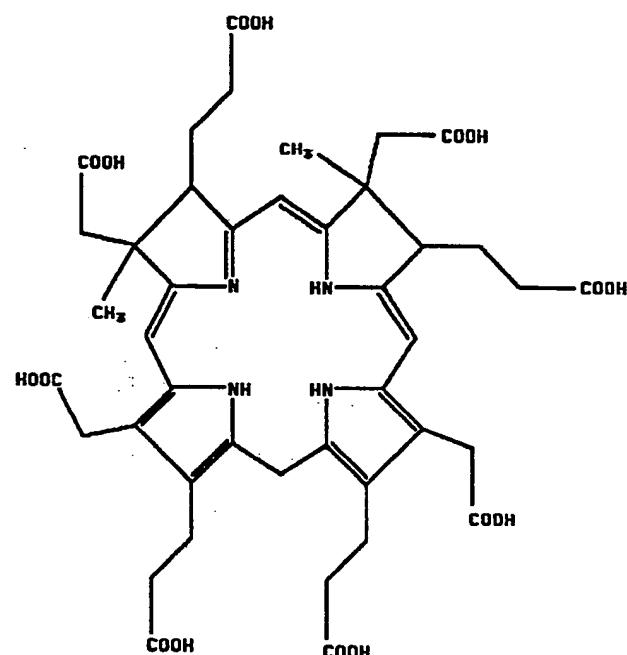


FIGURE 2

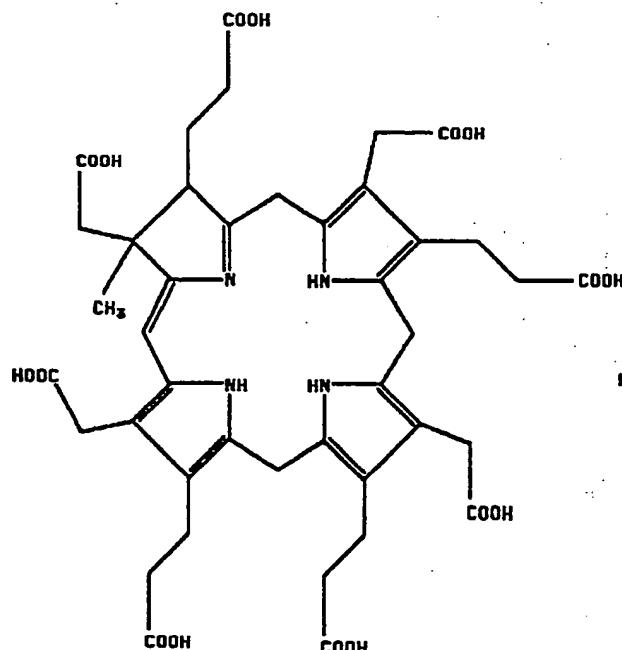
3/351



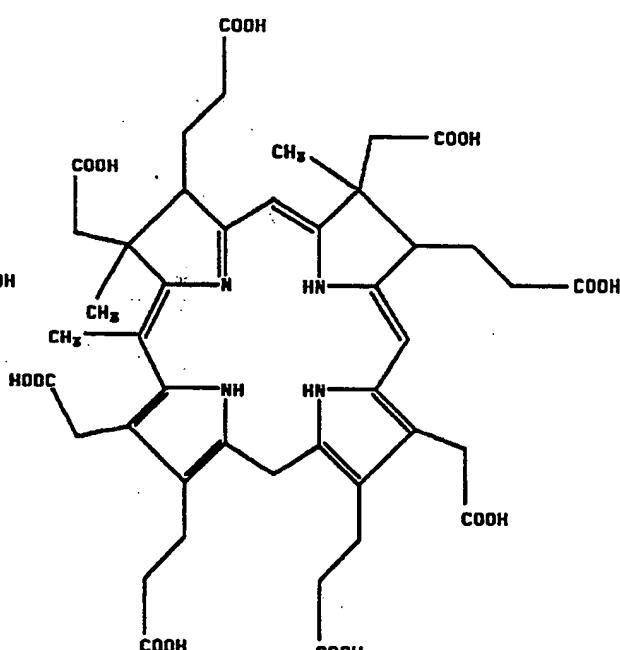
Uro'gen III



précorrine-2



précorrine-1



précorrine-3

FIGURE 3

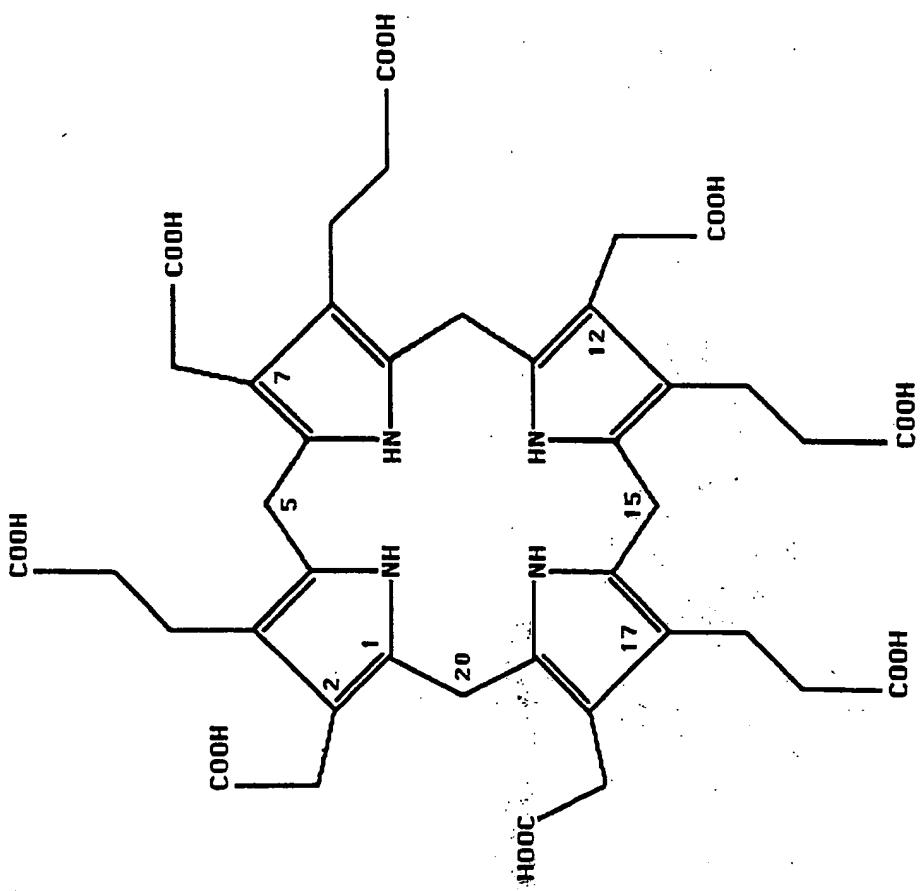
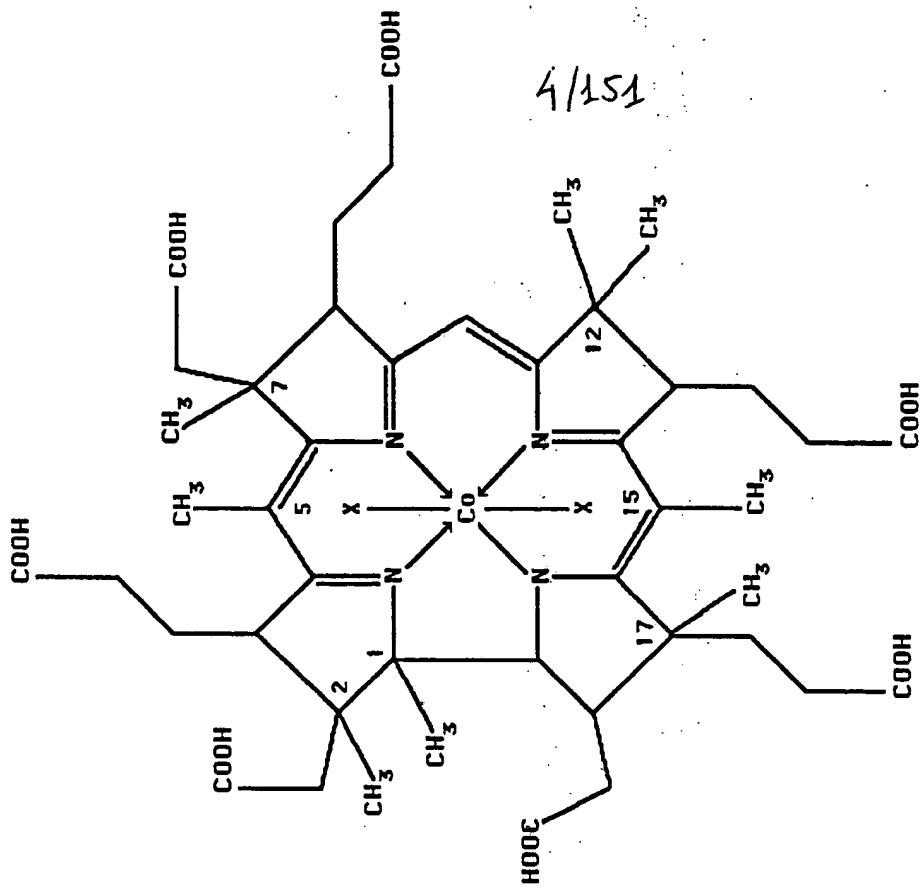


FIGURE 4

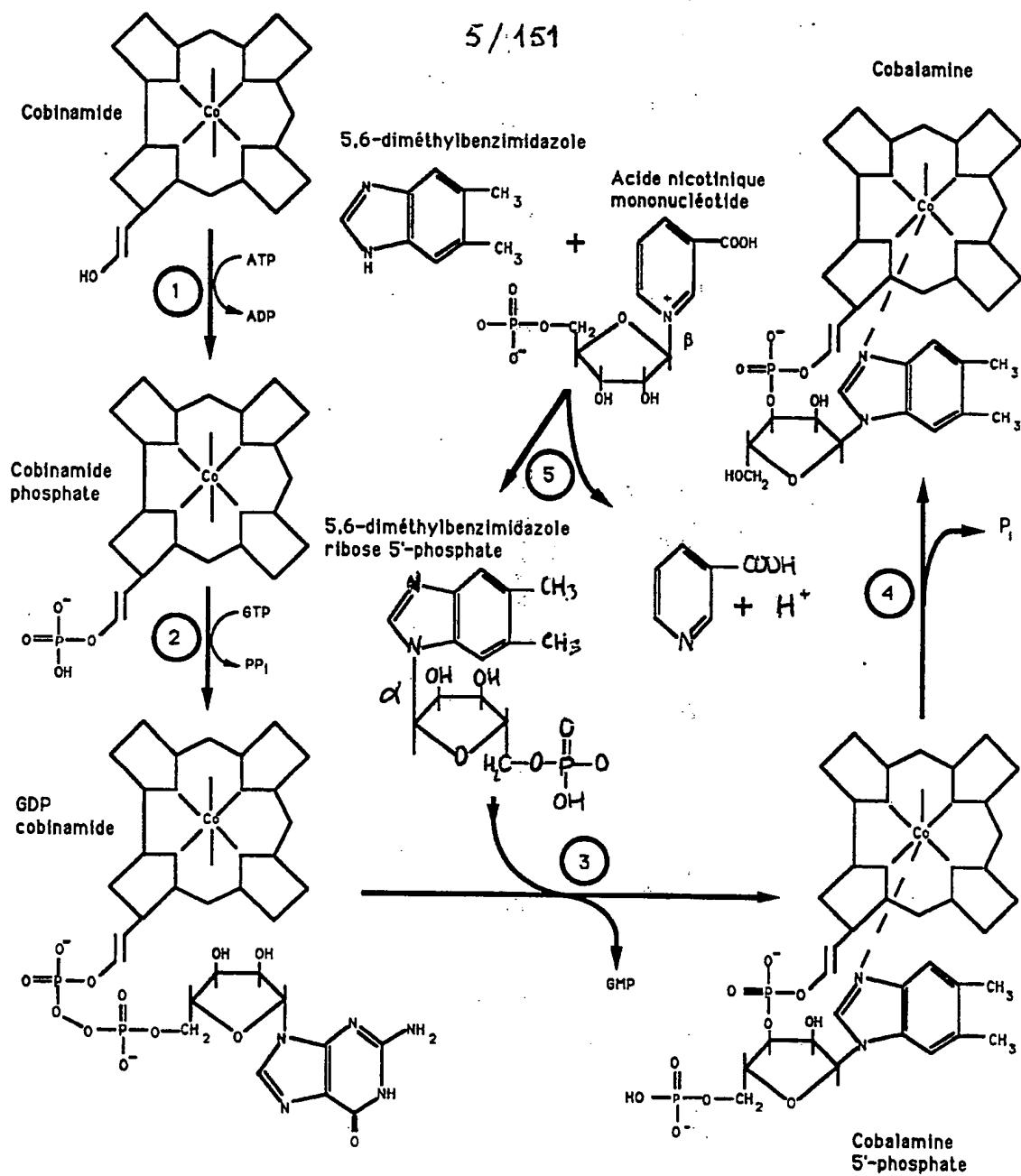


FIGURE 5

6 / 151

CARTE DE RESTRICTION DU FRANCHISEMENT CLAI-4-HINDU-HINDU-HINDU DE 5.4 KB

RESTRICTION DU FRAGMENT SALL-SALL-SALL-HOLI DE 4748 BP

卷之三

卷之三

CARTE

CARTE BE RESTRICTION BH FRAGMENT SET-SSET-1-Battall PE 3855 PB

0715 1750 2625 3500 4375 3250 6125 7000 7075. 0750

Figure 6.  
**FEUILLE DE REMPLACEMENT**

7 / 151

10            20            30            40            50            60  
 GGGCTGCAGG TCGACTCTAG AATCGATGAA GCCTGCGATG AAGGCAGGCGA CGAACAGGAA  
 CCCGACGTCC AGCTGAGATC TTAGCTACTT CGGACGCTAC TTCCGCCGCT GCTTGTCCCT  
  
 70            80            90            100          110          120  
 GGCGAGCAGG TGGAAGGCGA GATCTTGAC GGCAGGGACT CGAGAGGAGA GCTGTCAGGC  
 CCGCTCGTCC ACCTTCCGCT CTAGAACGTG CCGCCCCGTA GCTCTCCCT CGACAGTCCG  
  
 130          140          150          160          170          180  
 GGGATTTTCC GCCTTGTGTC AGAGCCCGGC GCGATTGCA AAGCCTTCTG TCGCGGTGTT  
 CCCTAAAAGG CGGAACACAG TCTCGGGCCG CGCTAACGT TTCGGAAGAC AGCGCCACAA  
  
 190          200          210          220          230          240  
 GCTGTCCATG CAGGTGTCGA AATTGAAAAA CCGACAAAGA TTCAAAGCCT TGTTCCAGCT  
 CGACAGGTAC GTCCACAGCT TAACTTTTT GGCTGTTTCT AAGTTTCGGA ACAAGGTCGA  
  
 250          260          270          280          290          300  
 CGCTGTCTTT CTGGATGGAG GCGCTCTCGC CCGCATGGTG CCGAAGAAGG GCTGTCCTTG  
 CGCACAGAAA GACCTACCTC CGCGAGAGCG GGCgtACCAC GGCTTCTITCC CGACAGGAAAC  
  
 310          320          330          340          350          360  
 CGATACGGTA GGCGGATGAC GATCTTCCTC AAACCGACA TGCGATGGC GCAATCCGGT  
 GCTATGCCAT CCGCCTACTG CTAGAAGGAG TTTGCCCTGT ACCGGCTACCG CGTTAGGCCA  
  
 370          380          390          400          410          420  
 TTGACCGGCC TTCCCGCGCTC CGGTAAAAAT GAAGGATATG CGACGGCGTC CGCTTTGGCG  
 AACTGGCCGG AAGGCGCGAG GCCATTTTA CTTCCATATAC GCTGCCGAG GCGAACCCGC  
  
 430          440          450          460          470          480  
 GACTGAAAGA GCGTCCGGTG CGGCCGACCC AGTCAGGGGG GCATCAGCCG GTGCTGTCCA  
 CTGACTTTCT CGCAGGCCAC GCGGCTGGG TCAGTCCCCC CGTAGTCGGC CACGACAGGT  
  
 490          500          510          520          530          540  
 GATCGGGCCGG GACGGATCGT CCCAGCCGGC GCTTCGTTAA GGAGAACAAAC GAAGGGAGCC  
 CTAGCCGGCC CTGCCTAGCA GGGTCGGCCG CGAAGCAATT CCTCTTGTG CTTCCCTCGG  
  
 550          560          570          580          590          600  
 GGCCGCCGAT GCCATCGGGC CAACACTCTG CACAGACGAC GAAAGCAGGA GCCGGGCTGG  
 CCGCGGGCTA CGGTAGCCCCG GTTGTGAGAC GTGCTGCTG CTTTCGTCCT CGGCCCCGACC  
  
 610          620          630          640          650          660  
 TGCTCGGGCT CGGCTGCGAG CGTCGACGCG CGGCCGAAGA GGTGATCGCC CTTGCCGAGC  
 ACGAGCCCCG ACGGACGCTC GCAGCGTGGC GCGGCTTCT CCACTAGCGG GAACGGCTCG  
  
 670          680          690          700          710          720  
 GTGCGCTTGC CGATGCCGGT GTTGCAGGGC GCGATCTGCG GCTGGTCGCC TCGCTCGATG  
 CACCGAACG GCTACGGCCA CAACCGGGGC CGCTAGACGC CGACCCAGCGG AGCGAGCTAC  
  
 730          740          750          760          770          780  
 CTCGCGCCGA GGAGCCGGCG ATCCCTGGCGG CCGCTGAGCA TTTCGCGTTT CCGGCCGCGT  
 GAGCGCGGCT CCTCGGCCGC TAGGACCGCC GGCAGTCGT AAAGCGCCAA GGCCGGCGCA  
  
 790          800          810          820          830          840  
 TCTACGATGC CGCCACGCTC GAAGCCGAAG CTTCCCGGCT CGCCAACCCG TCCGAGATCG  
 AGATGCTACG CGGGTGCAG CTTCCGCTTC GAAGGGCCGA GCGGTTGGGC AGGCTCTAGC  
  
 850          860          870          880          890          900  
 TCTTTGCCTA CACGGGTGT CATGGCGTG CCGAGGGTGC AGCGCTCGTC GGCGCCGGTC  
 AGAAACGGAT GTGCCAACAC GTACCGCAAC GGCTCCCACG TCGCGAGCAG CGCGGGCCAG  
  
 910          920          930          940          950          960  
 GCGAAGCCGT GCTGATTGTG CAGAAGATCG TCTCCGCCA TGCGACGGCC GCACTTGCCTG  
 CGCTTCGGCA CGACTAACAC GTCTTCTAGC AGAGGGGGT ACCTGCGCCGG CGTGAACGGC

FIGURE 7.1

8/151

970	980	990	1000	1010	1020
GGCCGGCGAC CTTGCGGCC GAAAAGCGCA TCCAGGCAGC GGAGGGCTGTC TGATGCATTC					
CCGGCCGCTG GAACGCGGG CTTCGCGT AGGTCCGCCG CCTCCGACAG ACTACGTAAG					
1030	1040	1050	1060	1070	1080
TTATGTTGT GAATTGAATC AATCTTTGC CCGGGGTTTC TCTCAAGTGG AATCCGGTTC					
AATAACAACAA CTTAACCTAG TTAGAAAACG GGCCCCAAAG AGAGTTCAACC TTAGGCCAAG					
1090	1100	1110	1120	1130	1140
TTTAGAGAGC GCGTCAGCGC TGCCGTTGGG TGGCGCGAA ATACAGGTGG GACAGCACGC					
AAATCTCTCG CGCAGTCCGC ACAGGCAACCC ACCGCGGCTT TATGTCCACC CTGTCGTGCG					
1150	1160	1170	1180	1190	1200
ATGATCGACG ACCTCTTGC CGGATTGCCG GCGCTCGAAA AAGGTTCGGT CTGGCTGGTC					
TACTAGCTGC TGGAGAAACG GCCTAACGGC CGCGAGCTT TTCCAAGCCA GACCGACCAAG					
1210	1220	1230	1240	1250	1260
GGCGCCGGCC CGGGCGATCC CGGCCTGTTG ACGCTGCATG CGGCAATGC GCTGGCCAG					
CCGCGGCCGG GGCGCTAGG GCCGGACAAC TGCGACGTAC GCCGGTTACG CGACGCGGTC					
1270	1280	1290	1300	1310	1320
GCGGATGTGA TCGTGCATGA TCGGCTGGTC AACGAGGATT GCCTGAAGCT CGCGCGGCCG					
CGCCTACACT AGCACGTACT ACGCGACCAAG TTGCTCTAA CGGACTTCGA GCGCGCCGGC					
1330	1340	1350	1360	1370	1380
GGCGCCGTGC TGGAGTTTCG GGGCAAGCGT GGCGGCAAGC CGTCGCCGAA GCAGCGCGAC					
CCGCGGCACG ACCTCAAACG CCCGTTCGA CCGCCGTTCG GCAGCGGCTT CGTCGCGCTG					
1390	1400	1410	1420	1430	1440
ATCTCGCTTC GCCTCGTCGA ACTCGCGCGC GCCGGCAACC GGGTGTGCG CCTCAAAGGC					
TAGAGCGAAG CGGAGCAGCT TGAGCGCGC CGGCCGTTGG CCCACGACGC GGAGTTCCG					
1450	1460	1470	1480	1490	1500
GGCGATCCCT TCGTCTTCGG TCGCGGTGGC GAGGAGGCGC TGACGCTGGT CGAACACCG					
CCGCTAGGGGA AGCAGAAGCC AGCGCCACCG CTCCCTCCGCG ACTGCGACCA GCTTGTGGTC					
1510	1520	1530	1540	1550	1560
GTGCCGTTCC GAATCGTGCC CGGCATCACC GCCGGTATCG GCGGGCTTGC CTATGCCGGC					
CACGGCAAGG CTTAGCACGG GCCGTAGTGG CGGCCATAGC CGCCCGAACG GATACGGCCG					
1570	1580	1590	1600	1610	1620
ATTCCCGTGA CCCATCGCGA GGTCAACCAC GCGGTCACTT TCCTGACTGG CCATGATTCC					
TAAGGGCACT GGGTAGCGCT CCAGTTGGTG CGCCAGTGAA AGGACTGACC GGTACTAAGG					
1630	1640	1650	1660	1670	1680
TCCGGCTGG TGCCGGATCG CATCAACTGG CAGGGCATCG CCAGCGGCTC GCCTGTCATC					
AGGCCGGACC ACGGCCTAGC GTAGTTGACC GTCCCCTAGC GGTGCGCGAG CGGACAGTAG					
1690	1700	1710	1720	1730	1740
GTCATGTACA TGGCGATGAA ACATATCGGC GCGATCACCG CCAACCTCAT TGCCGGCGC					
CAGTACATGT ACCGCTACTT TGTATAGCGC CGCTAGTGGC GGTGGAGTA ACGGCCGCCG					
1750	1760	1770	1780	1790	1800
CGCTCGCCGG ACGAACCGGT CGCCCTCGTC TGCAACGCCG CGACGCCGCA GCAGGGCGTG					
GCGAGCGGCC TGCTTGGCCA GCGGAAGCAG ACGTTGCCGC GCTGCCGCGT CGTCCGCCAC					
1810	1820	1830	1840	1850	1860
CTGGAAACGA CGCTTGCAGC TGCAGAGGCC GATGTTGCCG CGGCAGGGCT GGAGCCGCCG					
GACCTTTGCT CGAACCGCG ACGTCTCCGG CTACAAACGCC GCCGTCCCGA CCTCGCCGGC					
1870	1880	1890	1900	1910	1920
GCGATCGTCG TCGTCGGCGA GGTGGTGCAG CTGCGCGCAG CGCTCGACTG GATCGCCGCG					

9/151

CGCTAGCAGC AGCAGCCGCT CCACCACGCC GACGCGCGTC GCGAGCTGAC CTAGCCGC

1930	1940	1950	1960	1970	1980
CTGGACGGGC	GCAAGCTTGC	CGCCGACCCG	TTCGCCAATC	GCATTCTCAG	GAACCCGGCA
GACCTGCCCG	CGITCGAACG	GC GGCTGGGC	AAGCGGTTAG	CGTAAGAGTC	CTTGGGCCGT
1990	2000	2010	2020	2030	2040
TGAGCGGATT	GCTGATTGCC	GCACCCGCGT	CCGGCTCCGG	CAAGACGACG	GTGACGCTCG
ACTCGCCTAA	CGACTAACGG	CGTGGGCGCA	GGCCGAGGCC	GTTCTGCTGC	CACTGCGAGC
2050	2060	2070	2080	2090	2100
GGCTGATGCG	CGCCCTGAAG	AGGCGCGGCG	TGGCGATCGC	GCCCGGCAAG	GCGGGGCCGG
CCGACTACGC	CGGGGACTTC	TCCGCGCCGC	ACCGCTAGCG	CGGGCCGTTTC	CGCCCCGGCC
2110	2120	2130	2140	2150	2160
ACTATATCGA	TCCCCTTTC	CACGCGGCAG	CGACCCGCGA	GCCCTGCTTC	AACTACGACC
TGATATAGCT	AGGGCGAAAG	GTGCGCGTC	GCTGGCGCT	CGGGACGAAG	TTGATGCTGG
2170	2180	2190	2200	2210	2220
CCTGGCGAT	GCGCCCGGAA	CTGCTGCTTG	CCAATGCGTC	GCATGTGGCC	TCCGGCGGGC
GGACCCGCTA	CGCGGGCCTT	GACGACGAAC	GGTTACGCG	CGTACACCGG	AGGCCGCCCCG
2230	2240	2250	2260	2270	2280
GCACATTGAT	CGTCGAGGCG	ATGATGGGAC	TGCATGACGG	TGCTGCCGAC	GGCTCGGGAA
CGTGTAACTA	GCAGCTCCGC	TACTACCCCTG	ACGTACTGCC	ACGACGGCTG	CCGAGCCCTT
2290	2300	2310	2320	2330	2340
CGCCAGCGGA	CCTCGCCGCG	ACGCTGAACC	TTGCGGTCAT	TCTGGTGGTC	GATTGCGCCC
CGGGTCCGCT	GGAGCGGCAC	TGCGACTTGG	AACGCCAGTA	AGACCACCAAG	CTAACGCGGG
2350	2360	2370	2380	2390	2400
GCATGTCCCA	GTCGGTTGCC	GCCCTCGTGC	GCGGCTATGC	GGATCATCGC	GACGATATCC
CGTACAGGGT	CAGCCAACGG	CGGGAGCACG	CGCCGATACTG	CCTAGTAGCG	CTGCTATAGG
2410	2420	2430	2440	2450	2460
GGGTGGTTGG	CGTCATCCTC	AAACAAGGTG	GCAGCGATCG	GCATGAAATG	ATGCTGCGCG
CCCACCAACC	GCAGTAGGAG	TTGTTCCAGC	CGTCGCTAGC	CGTACTTTAC	TACGACGCGC
2470	2480	2490	2500	2510	2520
ATGCGCTCGG	CAAGGTGCGC	ATGCTGCT	TCGGCGTGT	CCGGCAGGAC	AGCGCATTGC
TACCGCAGCC	GTTCCACGCG	TACGGACAGA	AGCCGACCGA	GGCCGTCTG	TCGCGTAACG
2530	2540	2550	2560	2570	2580
AACTGCCGGA	GCGCCATCTC	GGGCTCGTC	AGGCAGGGCGA	ACACTCAGCG	CTTGAGGGCT
TTGACGGCCT	CGCGGTAGAG	CCCGAGCACG	TCCGCCCGCT	TGTGAGTCGC	GAACCTCCGA
2590	2600	2610	2620	2630	2640
TCATCGAGGC	GGCGGCCGCG	CGGGCTCGAGG	CTGCCCTGCGA	TCTCGACGCC	ATCCGCCCTGA
AGTAGCTCCG	CCGCCGGCGC	GCCCAGCTCC	GACGGACGCT	AGAGCTGCGG	TAGGCGGACT
2650	2660	2670	2680	2690	2700
TCGCGACGAT	TTTCCCGCAG	GTGCCCCGGG	CGGGCGATGC	CGAGCGTTTG	CGGCCGCTCG
AGCGCTGCTA	AAAGGGCGTC	CACGGGCGCC	GCGGCTACG	GCTCGCAAAC	GCCGGCGAGC
2710	2720	2730	2740	2750	2760
GTCAGCGCAT	CGCGGTGCGC	CGCGATATCG	CCTTTCGCTT	CTGCTACGAG	CACCTGCTTT
CAGTCGCGTA	GCGCCAGCGC	GCGCTATAGC	GGAAACGGAA	GACGATGCTC	GTGGACGAAA
2770	2780	2790	2800	2810	2820
ACGGCTGGCG	GCAAGGGCGC	GCGGAGATT	CCTTCTTCTC	GCCGCTCGCC	GACGAGGGGC
TGCCGACCGC	CGTCCGCGCG	CGCCTCTAAA	GGAAAGAAGAG	CGCGAGCGG	CTGCTCCCCG
2830	2840	2850	2860	2870	2880

FIGURE 7.3

10/151

CGGATGCGGC AGCCGATGCC GTCTATCTTC CGGGGGTTA TCCGGAGCTG CATGCGGGGC  
 GCCTACGCCG TCGGTACGG CAGATAGAAC GCCCCAAAT AGGCCTCGAC GTACGCCCG  
  
 2890 2900 2910 2920 2930 2940  
 AGCTGAGCGC CGCCGCCGA TTCCGTTCCG GCATGCATTG CGCGGCGGAA CGCGGCCCG  
 TCGACTCGCG CGGGGGCT AAGGCAAGGC CGTAGCTAAG GCGCCGCCTT GCGCCGCCGG  
  
 2950 2960 2970 2980 2990 3000  
 GCATCTTCGG CGAGTGCGGC GGCTATATGG TGCTCGCGA AGGGCTTGTG GCTGCCGATG  
 CGTAGAACCC GCTCACGCCG CCGATATAACC ACGAGGCCGT TCCCGAACAG CGACGGCTAC  
  
 3010 3020 3030 3040 3050 3060  
 GCACACGCTA CGACATGCTC GGCCTGCTGC CGCTCGTAAC CAGTTTGCC GAGGCCAGGC  
 CGTGTGCGAT GCTGTACGAG CGGACGACG GCGAGCATTG GTCAAAACGG CTCGGTCCG  
  
 3070 3080 3090 3100 3110 3120  
 GGCACCTCGG CTATCGCCGC GTCGTGCCGT TCGACAAACGC CTTCTTCGAT GGACCCATGA  
 CGGTGGAGCC GATAGCGGC GAGCACGGAC AGCTGTGCG GAAGAACGTA CCTGGGTACT  
  
 3130 3140 3150 3160 3170 3180  
 CGGCGCACGA ATTCCACTAT GCGACCATCG TCGCCGAAGG GGCGGCCGAT CGGCTGTTG  
 GCCCGCTGCT TAAGGTGATA CGCTGGTAGC AGCGGCTTCC CGGCGGGCTA CCCGACAAAC  
  
 3190 3200 3210 3220 3230 3240  
 CGGTCAAGCGA CGCCGCCGGC GAGGATCTCG GCCAGCGGG CCTCCGGCGC GGCCCTGTCG  
 CCCAGTCGCT CGGGCGGCCG CTCCTAGAGC CGTCCGGCCC GGAGGCCGG CCGGGACAGC  
  
 3250 3260 3270 3280 3290 3300  
 CGGGTTCCCT CATGCATCTG ATCGACGTG CAGGTGCTGC ATGAGGCCAC CGATCGTTCA  
 CCCAAGGAA GTACGTAGAC TAGCTGCAGC GTCCACGACG TACTCGCGTG GCTAGCAAGT  
  
 3310 3320 3330 3340 3350 3360  
 TGGTGGCGGC ATCACCGAGG CCGCAGCGG CTATGGCGGC CGGCCTGAAG ACTGGCTCGA  
 ACCACCGCCG TAGTGGCTCC GGCGTGCCTGC GATACCGCCG GCGGGACTTC TGACCGAGCT  
  
 3370 3380 3390 3400 3410 3420  
 TCTGTCGACC GGCATCAATC CATGCCCGT CGCGGTGCCG CGGGTCCCTG AGCGCCGCTG  
 AGACAGCTGG CGTAGTTAG GTACGGGCA GCGGAACGGG CGCCAGGGAC TCGCGCGGAC  
  
 3430 3440 3450 3460 3470 3480  
 GCACCGGCTG CCGGATCGGC AGACGGTAGA TGATGCGCGG AGCGCCGCCG CCGACTACTA  
 CGTGGCCGAC GGCCTAGCCG TCTGCCATCT ACTACCGGCC TCGCGCGGC GGCTGATGAT  
  
 3490 3500 3510 3520 3530 3540  
 CCGCACCAAC GGCCTGCTGC CTTTGGCGGT GCCGGSCACC CAGTCGGTGA TCCAGCTCCT  
 GGCGTGGTTG CGCACGACG GAAACGGCCA CGGCCCGTGG GTCAGCCACT AGGTGGAGGA  
  
 3550 3560 3570 3580 3590 3600  
 GCCACGTCTT GCTCCGGCCA ACAGGCACGT CGCGATTITC GGGCCGACCT ATGGCGAGTA  
 CGGTGCAGAA CGAGGCGGGT TGTCGTGCA GCGCTAAAAG CCCGGCTGGA TACCGCTCAT  
  
 3610 3620 3630 3640 3650 3660  
 TGCCCCGCTG CTTGAAGCGG CGGGCTTGC TGTGATCGC GTCGCGGATG CCGACCGCGT  
 ACGGGCGCAC GAACTTCGCC GGCGAAACG ACAGCTAGCG CAGCGCCTAC GGCTGCGCGA  
  
 3670 3680 3690 3700 3710 3720  
 CACGGCCGAA CATGGGCTTG TCATCGTGT CAACCCCAAC AACCCGACCG GCCGCCCTT  
 GTGCCGGCTT GTACCCGAAC AGTAGCAGCA GTTGGGGITG TTGGGCTGGC CGGCGCGGAA  
  
 3730 3740 3750 3760 3770 3780  
 GGCGCCGGCG GAGCTCTGG CGATCGCCGC AAGGCAGAAC GCGAGCGGG GACTGCTGCT  
 CGCGGGCCGC CTCGAAGACC GCTAGCGCGC TTCCGTCTTC CGCTCGCCGC CTGACGACGA

11/151

3790	3800	3810	3820	3830	3840
GGTCGATGAG	GCCTTCGGCG	ATCTTGAGCC	GCAACTGAGT	GTCGCTGGTC	ACGCGTCAGG
CCAGCTACTC	CGGAAGCCGC	TAGAACTCGG	CGTTGACTCA	CAGCGACCAAG	TGCGCAGTCC
3850	3860	3870	3880	3890	3900
GCAAGGCAAC	CTCATCGTCT	TCCGCTCCCT	CGGCAAGTTC	TTCGGCCATTG	CGGGCCTGCG
CGTTCCGTTG	GAGTAGCAGA	AGGCAGGAA	GCCGTTCAAG	AAGCCGGAAC	GCCC GGACGC
3910	3920	3930	3940	3950	3960
CCTCGGGCTTC	GTCGTTGCGA	CCGAGCCAGT	GCTTGGATCC	TTTGGCGATT	GGCTCGGTCC
GGAGCCGAAG	CAGCAACGCT	GGCTCGGTCA	CGAACGTAGG	AAA CGGCTAA	CCGAGCCAGG
3970	3980	3990	4000	4010	4020
CTGGGCTGTC	TCCGGCCCCGG	CGTTGACGAT	CTCGAAAGCG	CTGATGCAGG	GCGATACGAA
GACCCGACAG	AGGCCGGGGCC	GCAACTGCTA	GAGCTTTCGC	GACTACGTCC	CGCTATGCTT
4030	4040	4050	4060	4070	4080
GGCGATCGCG	GCGGGCATCC	TCGAGCGTCG	CGCCGGCCTC	GATGCGGCTC	TCGATGGGGC
CCGCTAGCGC	CGCCCGTAGG	AGCTCGCAGC	CGGGCCGGAG	CTACGCCGAG	AGCTACCCCG
4090	4100	4110	4120	4130	4140
AGGGCTAAC	CGTATCGGCG	GCACGGGGCT	ATTCTGCTG	GTGAGGATAC	CCAGGGCAGC
TCCCAGTTG	GCATAGCCGC	CGTCCCCCGA	TAAGCACGAC	CAGCTCGTAG	GGTCCCGTCG
4150	4160	4170	4180	4190	4200
TCTGCTGCAG	GAGCGGCTCT	GCGAGGCCA	TATTCTCACG	CGCAAGTTCG	ACTATGCCCG
AGACGACGTC	CTCGCCGAGA	CGCTCCGGGT	ATAAGAGTGC	GGTTCAAGC	TGATACGGGG
4210	4220	4230	4240	4250	4260
GACCTGGCTC	AGGGTCGGTC	TTGCGCCTGA	CGCGGCTGGT	GACCGACGGC	TGGCGGACGC
CTGGACCGAG	TCCCAGCCAG	AACGCGGACT	CGGCCGACCA	CTGGCTGCCG	ACCGCCTGCG
4270	4280	4290	4300	4310	4320
GCTTGGCCCGC	ATGGAGCTCT	GAGGTGTGCG	AGACGATCCT	GCTCATTCCTC	GCGCTGGCGC
CGAACGGCGC	TACCTCGAGA	CTCCACAGCC	TCTGCTAGGA	CGAGTAAGAG	CGCGACCGCG
4330	4340	4350	4360	4370	4380
TGGTGATCGA	CCGGCTTGTC	GGCGATCOGG	ACTGGCTCTG	GGCGCGCGTG	CCGCATCCGG
ACCACCTAGCT	GGCGCAACAG	CCGCTAGGCC	TGACCGAGAC	CCGCGCGCAC	GGCGTAGGCC
4390	4400	4410	4420	4430	4440
TCGTGTTTT	CGGCAAGGCC	ATCGGCTTT	TCGACCGCGCG	GCTGAACCGG	GAGGACCTCG
AGCACAAAAAA	GCCGTTCCGG	TAGCCGAAAAA	AGCTGCGCGC	CGACTTGGCC	CTCCTGGAGC
4450	4460	4470	4480	4490	4500
AGGATAGCGC	GGCGAAATTTC	CGTGGCGTCG	TCGCGATCCT	TTTGTGCTT	GGCATCAGCG
TCCTATCGCG	CGCGTTAAA	GCACCGCAGC	AGCGCTAGGA	AAACAAACGAA	CCGTAGTCGC
4510	4520	4530	4540	4550	4560
CCTGGTTCGG	CCATCTGCTG	CATCGCCTGT	TCGCCGTCT	CGGACCGCTC	GGCTTCTGC
GGACCAAGCC	GGTAGACGAC	GTAGCGGACA	AGCGGCAGGA	GCCTGGCGAG	CCGAAAGACG
4570	4580	4590	4600	4610	4620
TCGAGGGCGGT	TCTGGTCGCG	GTCTTCCCTGG	CACAGAAGAG	CCTCGCCGAT	CACGTGCGTC
AGCTCCGCCA	AGACCAAGCGC	CAGAAGGACC	GTGTCTTCTC	GGAGCGGCTA	GTGCACGCAG
4630	4640	4650	4660	4670	4680
GGCGTGGCCGG	GGGCTTGCGA	CAGGGCGGGC	TGGAAGGCAG	GGGTGCCGCC	GTGTCGATGA
CGCACCGGCC	CCCGAACGCT	GTCCCCGCCG	ACCTCCGCC	CGCACGGGG	CACAGCTACT
4690	4700	4710	4720	4730	4740
TCGTTGGTCG	CGATCCAAAG	ACGCTCGACG	AGCGGGCGGT	CTGCCGTGCC	GCGATCGAAA
AGCAACCAGC	GCTAGGTTTC	TGCGAGCTGC	TGGCGEGCCA	GACGGCACGG	CGCTAGCTTT

FIGURE 7.5

12/151

4750        4760        4770        4780        4790        4800  
 GCCTTGCCTGA GAATTTCTCC GACGGCGTCG TGGCGCCGGC CTTCTGGTAC GCGGTTGCCG  
 CGGAACGGCT CTTAAAGAGG CTGCCGCAGC ACCGCGGCCG GAAGACCATG CGCCAACGGC  
  
 4810        4820        4830        4840        4850        4860  
 GCCTGCCGGG GCCTCTTGCC TACAAGATGC TGAACACCGC CGATTGATG ATCGGCCACA  
 CGGACGGCCC CGAACGACGG ATGTTCTACG ACITGTGGCG GCTAAGCTAC TAGCCGGTGT  
  
 4870        4880        4890        4900        4910        4920  
 AGTCGCCGAA ATATCTGCAC TTGGCTGGG CCTCGGCCCG ACTCGACGAT CTCGCCAAC  
 TCAGCGGCTT TATAGACGTG AAGCCGACCC GGAGCCGGC TGAGCTGCTA GAGCGGTTGG  
  
 4930        4940        4950        4960        4970        4980  
 TGCCGGCAGC GAGGCTCTCG ATCCCTTTGA TCTCAGCCGG TGCGCTGATC CATCGTGGCG  
 ACGGCCGTCG CTCCGAGAGC TAGGAAAAGT AGAGTCGGCC AC CGCACTAG GTAGCACCGC  
  
 4990        5000        5010        5020        5030        5040  
 CCAGCGCCGC CAAGGATGCG CTGACCGTGG CCCTTCGCGA CCATGGCCTG CACCGCTCGC  
 GGTCGCGGGCG GTTCTTACGC GACTGGCACC GGGAAAGCGCT GGTACCGGAC GTGGCGAGCG  
  
 5050        5060        5070        5080        5090        5100  
 CGAACTCCGG CGGGCCATGG CCGGCGCGCT CGATCTGCAG CTTGCCGGTC  
 GCTTGAGGCC GACCGGCCTT CGCCGGTACC GGCGCGCGA GCTAGACGTC GAACGGCCAG  
  
 5110        5120        5130        5140        5150        5160  
 CGCGGATCTA TGGCGCGTC AAGGTCAGCG AACCTATGAT CAACGGTCCG GGCGGAGCGG  
 GCGCCTAGAT ACCGGCCGAG TTCCAGTCGC TTGGATACTA GTTGCCAGGC CGGGCTCGCC  
  
 5170        5180        5190        5200        5210        5220  
 TTGCAACAAAG CGAAGACATC GACGCCGTA TTGCTGTATT TTATGGCGCC TGTACGGTCA  
 AACGTTGTTG AACCTCTGTAG CTGCGGCCAT AACGACATAA ATAACCGGG ACATGCCAGT  
  
 5230        5240        5250        5260        5270        5280  
 TGGCCGGGTT TGTTCTTGCA ATCGCAATGA TTTGATCGCG GAAGTTGACC TTCGCATTAA  
 ACCGGCCCAA ACAAGAACGT TAGCGTTACT AAACTAGCGC CTTCAACTGG AAGCGTAATT  
  
 5290        5300        5310        5320        5330        5340  
 GACTCTGCTT TCCATATGTA TTAAGATCGT ATCATATTG ATCAGTTATT CTCCCTGGAAC  
 CTGAGACGAA AGGTATACAT AATTCTAGCA TAGTATAAGC TAGTCAATAA GAGGACCTTG  
  
 5350        5360        5370        5380        5390        5400  
 GTTGGTTCC ACCGGTACGT GTTCGTCTTC CCGGAGAGAG AAGCATGCGC AAAAGCTT  
 CAAACCAAGG TGGCCATGCA CAAGCAGAAG GGCCTCTCTC TTCTGTACGCG TTTTCGAA

FIGURE 7.6

13/151

10	20	30	40	50	60
GAATTGCCA	GCGCCTACAT	GGCTGACCTC	AAGCAGTTC	TCCGTGGCCC	GAAGAACGAG
CTTAAGCGGT	CGCGGATGTA	CCGACTGGAG	TTCGTCAGG	AGCACCGGGGT	CTTCTTGCTC
70	80	90	100	110	120
GGCCGGGAGA	TTTTCCCTCG	CGGGCCTGAG	TATTTTCGCG	CGCTCGACCT	GACGCCGCTC
CCGGCCGTCT	AAAAGGGAGC	GCCCCGACTC	ATAAAAGCGC	GCGAGCTGGA	CTGCGCGAG
130	140	150	160	170	180
GACAAGGTGC	GCGTGGTCAT	TCTCGGCCAG	GATCCCTATC	ACGGTACGCG	CCAGGGCGCAT
CTGTTCCACG	CGCACCAAGTA	AGAGCCGGTC	CTAGGGATAG	TGCCACTGCC	GGTCCCGCTA
190	200	210	220	230	240
GGGCTCTGCT	TCAGCGTTCG	CCCCGGTGTG	CGGACGCCGC	CGTCGCTGGT	CAACATCTAC
CCCGAGACGA	AGTCGCAAGC	GGGGCCACAG	GCCTGCGCG	GCAGCGACCA	GTTGTAGATG
250	260	270	280	290	300
AAGGAACCTGA	ATACCGATCT	CGGTATTCCG	CCGGCGCGTC	ACGGTTTTCT	CGAAAGCTGG
TTCCTTGACT	TATGGCTAGA	GCCATAAGGC	GGCCCGCGAG	TGCCAAAAGA	GCTTCGACC
310	320	330	340	350	360
GCAAGGCAGG	GCGTGTGCT	TTTGAACAGC	GTGCTGACGG	TAGAGCGCGG	GAACGTGCGT
CGTTCCGTCC	CGCACGACGA	AAACTTGTG	CACGACTGCC	ATCTCGCGCC	CTTGCACGCA
370	380	390	400	410	420
CACACCAGGG	TCACGGTTGG	GAAAAGTTCA	CGGATGCGAT	CATCCGTGCG	GTCAACGAGG
GTGTGGTCCC	AGTGCCAACC	CTTTCAAGT	GCCTACGCTA	GTAGGCACGC	CAGTTGCTCC
430	440	450	460	470	480
CCGAGCATCC	CGTCGTCTTC	ATGCTTTGGG	GCTCTATGC	GCAGAAGAAG	GCGGCCTTCG
GGCTCGTAGG	GCAGCAGAACG	TACGAAACCC	CGAGGATACG	CGTCTTCTTC	CGCCGGAAGC
490	500	510	520	530	540
TCGACCGCTC	GCGCCATCTT	GTCCTGAGGG	CACCACATCC	GTGCGCGCTC	TCAGCCATT
AGCTGGCGAG	CGGGTAGAA	CAGGACTCCC	GTGGTGTAGG	CAGCGCGAG	AGTCGGTAA
550	560	570	580	590	600
CCGGCTTTCT	CGGCTGCCGG	CATTTTCCC	AGGCCAATGC	CTTCCTCGAA	AGCAAAGGCT
GGCCGAAAGA	GCCGACGGCC	GTAAAAAGGG	TCCGGTTACG	GAAGGAGCTT	TCGTTCCGA
610	620	630	640	650	660
TCGATCCGAT	CGACTGGCGG	CTGCGGGAAA	ATCCGGCTGC	GGACATCAAC	TGAAGGCTTG
AGCTAGGCTA	GCTGACCGCC	GACGGCCTTT	TAGGCCGACG	CCTGTAGTIG	ACTTCCGAAC
670	680	690	700	710	720
GCGCGAATGA	CGGCTTGTC	GTCGCCCTGA	GGTCTTGCCT	TGGCGCGGGC	GATCCGCCTA
CGCGCTTACT	GCCGAAACAG	CAGCGGGACT	CCAGAACCGA	ACCGCCGCCG	CTAGGCGGAT
730	740	750	760	770	780
AGACGCCCGA	ACGAAATGGC	GGAGGGGGC	ATGCGAAAAA	TTCTGATCAT	CGGCATCGGT
TCTGCGGGCT	TGCTTTACCG	CCTCCGCCCG	TACGCTTTT	AAGACTAGTA	GCCGTAGCCA
790	800	810	820	830	840
TCGGGCAATC	CCGAACACAT	GACCGTGAG	GCGATCAACG	CGCTGAACGTG	CGCCGACGTG
AGCCCCTTAG	GGCTTGTGTA	CTGGCACGTC	CGCTAGTTC	GCGACTTGCAC	CGGGCTGCA
850	860	870	880	890	900
CTCTTTATCC	CGACCAAGGG	AGCGAAGAAG	ACCGAGCTTG	CCGAAGTGC	CCGCGACATC
GAGAAATAGG	GCTGGTTCCC	TCGCTTCTTC	TGGCTCGAAC	GGCTTCACGC	GGCGCTGTAG
910	920	930	940	950	960
TGCGCCCGCT	ACGTCACGCG	CAAGGACAGC	CGCACCGTCG	AGTTCGCGGT	GCCCGTGC
ACGCGGGCGA	TGCAGTGC	GTTCTGTG	CGTGGCAGC	TCAAGCGCCA	CGGGCACGCC

FIGURE 8.1

14/151

970	980	990	1000	1010	1020
CGCACCGAAG	GCGTCAGCTA	TGACGGCAGC	GTCGATGACT	GGCACGCCA	GATCGCTGGG
CGGTGGCTTC	CGCAGTCGAT	ACTGCCGTCG	CAGCTACTGA	CCGTGCGGGT	CTAGCGACCC
1030	1040	1050	1060	1070	1080
ATTTACGAAG	CGCTTCTATC	GAAGGAGTTG	GGCGAAGAGG	GAACCTGGCGC	GTTTCTCGTC
TAAATGCTTC	GCGAAGATAG	CTTCCTCAAC	CCGCTCTCC	CTTGACCGCG	CAAAGAGCAG
1090	1100	1110	1120	1130	1140
TGGGGCGACC	CGATGCTCTA	TGACAGCACC	ATTGCATCG	TCGAGCGGGT	CAAGGCACGC
ACCCCGCTGG	GCTACGAGAT	ACTGTCGTGG	TAAGCGTAGC	AGCTCGCCCA	GTTCCGTGCG
1150	1160	1170	1180	1190	1200
GGTGAGGTCG	CCTTCGCCCTA	CGACGTCATT	CCCAGGATCA	CCAGTCTGCA	GGCGCTTGC
CCACTCCAGC	GGAAGCGGAT	GCTGCACTAA	GGGCCTAGT	GGTCAGACGT	CCGCGAAACG
1210	1220	1230	1240	1250	1260
GCCAGCCACC	GCATTCCGCT	GAACCTCGTC	GGCAAGCCGG	TGGAGATCAC	CACGGGGCGT
CGGTCGGTGG	CGTAAGGCGA	CTTGGAGCAG	CCGTTGGCC	ACCTCTAGTG	GTGCCCCGCA
1270	1280	1290	1300	1310	1320
CGGCTGCACG	AAAGCTTTC	CGAGAACGAGC	CAGACCTCGG	TCGTCATGCT	CGATGGCGAA
GCCGACGTGC	TTTCGAAAGG	GCTCTTCCTG	GTCTGGAGCC	AGCAGTACGA	GCTACCGCTT
1330	1340	1350	1360	1370	1380
CAGGCCTTTC	AGCGGGTCGA	GGACCCGGAG	GCGGAGATCT	ATTGGGGCGC	CTATCTCGGC
GTCCGCAAAG	TCGCCCCAGCT	CCTGGGCCCTC	CGCCTCTAGA	TAACCCCCGCG	GATAGAGCCG
1390	1400	1410	1420	1430	1440
ACGCGGGATG	AGATCGTCAT	TTCCGGCCGC	GTGGCTGAGG	TGAAGGACCG	GATCCTTGAA
TGCGCCCTAC	TCTAGCAGTA	AAGGCCGGCG	CACCGACTCC	ACTTCCTGGC	CTAGGAACCT
1450	1460	1470	1480	1490	1500
ACGCGGGCGG	CGGCGCGCGC	GAAGATGGGA	TGGATCATGG	ACATCTATCT	CCTGCGCAAG
TGCGCCCGCC	GCCGCGCGCG	CTTCTACCCCT	ACCTAGTACC	TGTAGATAGA	GGACCGGTT
1510	1520	1530	1540	1550	1560
GGCGCCGACT	TCGACGAGTG	ACGGGGAGGG	CCGATCTCGC	TCGTGTTGA	TCTCACTCAA
CCGCGGCTGA	AGCTGCTCAC	TGCCCCCTCCC	GGCTAGACGC	AGCACAAACT	AGAGTGAGTT
1570	1580	1590	1600	1610	1620
GGTTTGCGGC	TGTGTTATAG	CGTCTTAAGA	GGCTTCTTC	GGGAGGGAGAA	CCTCAAGTGA
CCAAACGCCG	ACACAATATC	GCAGAAATTCT	CCGAAGAACT	CCCTCCTCTT	GGAGTTCACT
1630	1640	1650	1660	1670	1680
TGACGGATTT	GATGACCAAGC	TGCGCCCTTC	CATTGACCGG	AGATGCCGGC	ACCGTCGCTT
ACTGCCTAAA	CTACTGGTCG	ACGCGGGAAAG	GTAACTGGCC	TCTACGGCCG	TGGCAGCGAA
1690	1700	1710	1720	1730	1740
CGATGCCCG	CGGCGCCTGC	CCGTCTTGG	CAGAGCCGAT	GCAGACCGGC	GACGGCCTGC
GCTACGCCGC	GCCGCGGACG	GGCAGGAACC	GTCTCGGCTA	CGTCTGGCCG	CTGCGGGACG
1750	1760	1770	1780	1790	1800
TCGTGAGGGT	GAGGCCAACG	GATGACAGCC	TGACGCTGCC	GAAGGTCATT	GCCCTTGCCA
AGCACTCCCA	CTCCGGTTGC	CTACTGTCGG	ACTGCGACGG	CTTCCAGTAA	CGGGAACGGT
1810	1820	1830	1840	1850	1860
CGGCTGCCGA	GCGCTTCGGC	AATGGCATCA	TCGAGATTAC	CGCGCGCGGA	AACCTGCAGC
GCCGACGGCT	CGCGAAGCCG	TTACCGTAGT	AGCTCTAATG	GCAGCGCGCT	TTGGACGTCG
1870	1880	1890	1900	1910	1920
TTCGCGGCCT	GAGCGCGGCT	TCGGTGCCAA	GGCTGGCGCA	GGCGATCGGC	GATGCGGAGA

15/151

AAGCGCCGGA CTCGCGCCGA AGCCACGGTT CCGACCAGGT CGCCTAGCCG CTACGCCCTCT  
 1930 1940 1950 1960 1970 1980  
 TCGCCATTGC CGAGGGGCTC GCGATCGAGG TGCCGCCCCCT GGCCGGCATC GACCCGGACG  
 AGCGGTAACG GCTCCCCGAG CGCTAGCTCC ACGGCGGGGA CGGGCCGTAG CTGGGCCCTGC  
 1990 2000 2010 2020 2030 2040  
 AGATCGCCGA TCCGCGGCCG ATTGCCACTG AGCTTCGTGA AGCGTTGGAT GTGCGCCAGG  
 TCTAGCGGCT AGGCGCCGGC TAACGGTACG TCGAAGCACTA CACGCGGTCC  
 2050 2060 2070 2080 2090 2100  
 TGCCGTTGAA GCTTGCACCC AAATTATCCG TCGTCATCGA TAGCGGTGGC CGGTTGGTC  
 ACGGCAACTT CGAACGTGGG TTTAATAGGC AGCAGTAGCT ATCGCEACCG GCCAAACCAG  
 2110 2120 2130 2140 2150 2160  
 TCGGGCGCTGT CGTCGCCGAC ATTCGCCTTC AGGCGGTTTC GACTGTGCG GGGGTGGCCT  
 AGCCGCGACA GCAGCGGCTG TAAGCGGAAG TCCGCAAAG CTGACAGCGC CCCCACCGGA  
 2170 2180 2190 2200 2210 2220  
 GGGTGCTGTC GCTTGGCGGC ACGTCAACGA AGGCATCGAG CGTCGGGACG TTGGCCGGCA  
 CCCACGACAG CGAACCGCCG TGCAGTTGCT TCCGTAGCTC GCAGCCCTGC AACCGGCCGT  
 2230 2240 2250 2260 2270 2280  
 ACGCGGTCGT GCCGGCCCTG ATCACCATTC TCGAGAAACT GGCGAGCCTG GGCACGACGA  
 TGCGCCAGCA CGGCCGGGAC TAGTGGTAAG AGCTCTTGA CCGCTCGGAC CCGTGCTGCT  
 2290 2300 2310 2320 2330 2340  
 TGCGCGGGCG CGATCTGGAC CCGTCGGAAA TCCGCGCGCT CTGTCGCTGT GAGACATCGT  
 ACGCGCCCGC GCTAGACCTG GGCAGCCITT AGGCGCGCGA GACAGCGACA CTCTGTAGCA  
 2350 2360 2370 2380 2390 2400  
 CCGAACGCCG GGCCGCTCCG CGTTGGCCG CAATACCCGG CATTATGCG CTGGGTAACG  
 GGCTTGCAGG CGGGCGAGGC GCAAGCCGGC GTTATGGGCC GTAAGTACGC GACCCATTGC  
 2410 2420 2430 2440 2450 2460  
 CCGACACCGT TCTCGGCCTC GGTCTGGCCT TTGCTCAGGT GGAGGCCGCC GCGCTGGCAT  
 GGCTGTGGCA AGAGCCGGAG CCAGACCGGA AACGAGTCCA CCTCCGGCGG CGCGACCGTA  
 2470 2480 2490 2500 2510 2520  
 CCTACCTGCA TCAGGTCCAG GCGCTTGGCG CCAATGCGAT CCGGCTTGCG CCCGGGCACG  
 GGATGGACGT AGTCCAGGTC CGCGAACCGC GGTTACGCTA GGCCGAACGC GGGCCCGTGC  
 2530 2540 2550 2560 2570 2580  
 CCTTCTTCGT CCTCGGCCTT TGCCCCGAGA CCGCGGCTGT GGCGCAGAGC CTGGCAGCGT  
 GGAAGAACGA GGAGCCGGAA ACGGGGCTCT GGCGCCGACA CGCGTCTCG GACCGTCGCA  
 2590 2600 2610 2620 2630 2640  
 CACACGGTTT TCGCATTGCC GAGCAGGATC CGCGCAATGC GATGCCACC TGCGCCGGCA  
 GTGTGCCAAA AGCGTAACGG CTCGTCCTAG GCGCGTTACG CTAGCGGTGG ACGCGGCCGT  
 2650 2660 2670 2680 2690 2700  
 GCAAGGGTTG CGCCTCGGCG TGGATGGAAA CCAAGGGCAT GGCGAGCGC CTCGTGAGA  
 CGTTCCCAAC CGGGAGCCGC ACCTACCTT GGTTCCCGTA CGCGTCTCG GAGCAGCTCT  
 2710 2720 2730 2740 2750 2760  
 CGGCGCCGGGA ATTGCTCGAC GGGTCGCTCA CGGTGCACTCT CTCCGGCTGC GCCAAGGGCT  
 GCCGCGGGCCT TAACCGAGCTG CCCAGCGAGT GGCACGTAGA GAGGCCGACG CGGTTCCCGA  
 2770 2780 2790 2800 2810 2820  
 GCGCCCGGCC GAAGCCGTCC GAACTGACGC TTGTCGGTGC GCCATCAGGA TACGGGCTTG  
 CGCGGGCCGG CTTCGGCAGG CTTGACTGCG AACAGCCACG CGGTAGTCCT ATGCCCGAAC  
 2830 2840 2850 2860 2870 2880

FIGURE 8.3

16/151

TCGTAAATGG GGCTGCCAAT GGCTTGCCAA GCGCCTACAC CGATGAGAAT GGAATGGGAT  
 AGCATTTACC CCGACGGTTA CCGAACGGTT CGCGGATGTG GCTACTCTTA CCTTACCCCTA  
  
 2890 2900 2910 2920 2930 2940  
 CCGCCCTTGC CCGGCTCGGC CGGCTGGTGC GGCAAAACAA AGACGCTGGC GAATCGGCGC  
 GGCGGGAAACG GGCGAGCCG GCCGACCAACG CCGTTTGTT TCTCGGACCG CTTAGCCGCG  
  
 2950 2960 2970 2980 2990 3000  
 AGTCCTGTCT TACACGGCTC GGAGCTGCAC GCCTCTCGGC AGCGTTCGAA CAGGGATAGA  
 TCAGGACAGA ATGTGCGGAG CCTCGACCGC CGCAGAGCCG TCGCAAGCTT GTCCCTATCT  
  
 3010 3020 3030 3040 3050 3060  
 CATGCCCTGAG TATGATTACA TTGCGATGG CAACGCCATC TACGAGCGTT CCTTCGCCAT  
 GTACGGACTC ATACTAATGT AAGCGCTTACG GTTGCCTGAG ATGCTCGCAA GGAAGCGGTA  
  
 3070 3080 3090 3100 3110 3120  
 CATCCCGGCC GAGGCCGATC TGTCGCGCTT CTCCGAAGAG GAAGCGGATC TGGCTGTGCG  
 GTAGGCCGG CTCCGGCTAG ACAGCGCGAA GAGGCCTCTC CTTCGCTAG ACCGACACGC  
  
 3130 3140 3150 3160 3170 3180  
 CATGGTGCAC GCCTCGGGTT CGCTCGAGGC GACCAGGCAG TTCGTGTTT CTCCCGATTT  
 GTACCACGTG CGGACGCCAA GGCACTTCCG CTGGTCCGTC AAGCACAAAA GAGGGCTAAA  
  
 3190 3200 3210 3220 3230 3240  
 CGTAAGCTCG GCCCGTGCAG CGCTGAAAGC CGGTGCGCCG ATCCTCTGCG ATGCCGAGAT  
 GCATTCGAGC CGGGCACGCC CGCACTTTCG GCCACCGCC TAGGAGACGC TACGGCTCTA  
  
 3250 3260 3270 3280 3290 3300  
 GGTTGCGCAC GGTGTCACCC GCGCCCGTCT GCCGGCCGGC AACGAGGTGA TCTGCACGCT  
 CCAACCGCTG CCACAGTGGG CGCGGGCAGA CGGCCGGCCG TTGCTCCACT AGACGTGCGA  
  
 3310 3320 3330 3340 3350 3360  
 GCGCGATCCT CGCACGCCCG CACTTGGCCG CGAGATCGGC AACACCCGCT CCGCCGCAGC  
 CGCGCTAGGA GCGTGCGGGC GTGAACGCCG GCTCTAGCCG TTGTGGCGA GGCAGCGCTCG  
  
 3370 3380 3390 3400 3410 3420  
 CCTGAAGCTC TGGAGCGAGC GGCTGGCCGG TTCGGTGGTC GCGATCGGCA ACGCGCCGAC  
 GGACTTCGAG ACCTCGCTCG CGCACCGCC AAGCCACCAAG CGCTAGCCGT TGCAGGGCTG  
  
 3430 3440 3450 3460 3470 3480  
 GGC GTT GTC TT C C T T G G AA AT G C T G C G CG AC GG CG CG C C G C G A A T C C T  
 CCGCAACAAAG AAGGAGAACCT TTACGACGC GCTGCCGCGC GGCTTCGGCC GCGTTAGGA  
  
 3490 3500 3510 3520 3530 3540  
 CGGCATGCC C G C G G T T T C G T C G G G C G G A A T C G A A G G A T G C T G G C C G A A C T C  
 GCGTACGGG CAGCAAAGC AGCCACGCC CGCTAGCTTC CTACGCGACC GGCTCTTGAG  
  
 3550 3560 3570 3580 3590 3600  
 CTATGGCGTT CCCCTCGCCA TCGTGCACGG CCGCCTCGGC GGGAGTGCCTA TGACGGCGGC  
 GATACCGCAA GGGAAAGCGGT AGCACCGCC CGCGGAGCCG CCCTCACGGT ACTGCCGCCG  
  
 3610 3620 3630 3640 3650 3660  
 AGCGCTTAAC TCGCTCGCGA GGCGGGCCCT GTGAGCGGGC TCGGCGTGGG GCGCCTGATC  
 TCGCGAATTG AGCGAGCGCT CGGGCCCCGA CACTCGCCGC AGCCGCACCC CGCGGACTAG  
  
 3670 3680 3690 3700 3710 3720  
 GGTGTTGGGA CGGGCCCCGG TGATCCGAA CTTTGACGG TCAAGGCGGT GAAGGCGCTC  
 CCACAACCCCT GGCGGGGCC ACTAGGCCCT GAAAAGGCC AGTTCCGCCA CTTCCGCGAG  
  
 3730 3740 3750 3760 3770 3780  
 GGGCAAGCCG ATGTGCTTGC CTATTTCGCC AAGGCCGGC GAAGCGGTAA CGGCCCGCG  
 CCCGTTCGGC TACACGAACG GATAAAGCGG TTCCGGCCCG CTTCGCCATT GCGGGCGCG

FIGURE 8.4

17/181

3790	3800	3810	3820	3830	3840
GTGGTCGAGG	GTCTGCTGAA	GCCCCATCTT	GTCGAGCTGC	CGCTATACTA	TCCGGTGACG
CACCAGCTCC	CAGACGACTT	CGGGCTAGAA	CAGCTCGACG	GGCATATGAT	AGGCCACTGC
3850	3860	3870	3880	3890	3900
ACCGAAATCG	ACAAGGACGA	TGGCGCCTAC	AAGACCCAGA	TCACCGACTT	CTACAATGCG
TGGCTTTAGC	TGTTCCCTGCT	ACCGCGGATG	TTCTGGGTCT	AGTGGCTGAA	GATGTTACGC
3910	3920	3930	3940	3950	3960
TCGGCCGAAG	CGGTAGCGGC	GCATCTTGCC	GCCGGGCGCA	CGGTGCGCGT	GTCAGTGAA
AGCCGGCTTC	GCCATCGCCG	CGTAGAACGG	CGGCCCCGCGT	GCCAGCGGCA	CGAGTCACCTT
3970	3980	3990	4000	4010	4020
GGCGACCCGC	TGTTCTATGG	TTCCTACATG	CATCTGCATG	TGCGGCTCGC	CAATCGTTTC
CCGCTGGCGC	ACAAGATACC	AAGGATGTAC	GTAGACGTAC	ACGCCGAGCG	GTTAGCAAAG
4030	4040	4050	4060	4070	4080
CCGGTCGAGG	TGATCCCCGG	CATTACCGCC	ATGTCCGGCT	GTTGGTCGCT	TGCCGGCCTG
GGCCAGCTCC	ACTAGGGGCC	GTAATGGCGG	TACAGGCCGA	CAACCAGCGA	ACGGCCGGAC
4090	4100	4110	4120	4130	4140
CCGCTGGTGC	AGGGCGACGA	CGTGCTCTCG	GTGCTTCCGG	GCACCATGGC	CGAGGCCGAG
GGCGACCACG	TCCCCTGCT	GCACGAGAGC	CACGAAGGCC	CGTGGTACCG	GCTCCGGCTC
4150	4160	4170	4180	4190	4200
CTCGGCCGCA	GGCTTGCGGA	TACCGAACCC	GCCGTGATCA	TGAAGGTCGG	GCGCAATTG
GAGCCGGCGT	CCGAACGCCT	ATGGCTTCGG	CGGCACTAGT	ACTTCCAGCC	CGCGTTAAC
4210	4220	4230	4240	4250	4260
CCGAAGATCC	GTCGGGCGCT	CGCTGCCCTCC	GGCCGCTCTCG	ACCAGGCCGT	CTATGTCGAA
GGCTTCTAGG	CAGCCCGCGA	GCGACGGAGG	CCGGCAAGAGC	TGGTCCGGCA	GATACAGCTT
4270	4280	4290	4300	4310	4320
CGCGGCCACGA	TGAAGAACCG	GGCGATGACG	GCTCTTGC GG	AAAAGGCCGA	CGACGAGGCG
GCGCCGTGCT	ACTTCTTGCG	CCGCTACTGC	CGAGAACGCC	TTTCCGGCT	GCTGCTCCGC
4330	4340	4350	4360	4370	4380
CCCTATTTCT	CGCTGGTGCT	CGTCCCCGGC	TGGAAGGACC	GACCATGACC	GGTACGCTCT
GGGATAAAGA	GCGACCACGA	GCAAGGGCCG	ACCTTCTGG	CTGGTACTGG	CCATGCGAGA
4390	4400	4410	4420	4430	4440
ATGTCGTCGG	TACCGGACCG	GGCAGCGCCA	AGCAGATGAC	GCCGGAAACG	GCGGAAGCCG
TACAGCAGCC	ATGGCCTGGC	CCGTCGCGGT	TCGCTACTG	GGGCTTTGC	CGCCTTCGGC
4450	4460	4470	4480	4490	4500
TTGCGGCCGC	TCAGGAGTTT	TACGGCTACT	TTCCCTATCT	CGACCGGCTG	AACCTCAGAC
AACGCCGGCG	AGTCCTCAAA	ATGCCGATGA	AAGGGATAGA	GCTGGCCGAC	TTGGAGTCTG
4510	4520	4530	4540	4550	4560
CGGATCAGAT	CCGTGTCGCC	TCGGACAACC	GCGAGGGAGCT	CGATCGGGCA	CAGGTCGCGC
GCCTAGTCTA	GGCACAGCGG	AGCCTGTTGG	CGCTCCCTCGA	GTAGCCCGT	GTCCAGCGCG
4570	4580	4590	4600	4610	4620
TGACGCGGGC	TGCGGGCAGGC	GTGAAGGTCT	GCATGGTCTC	CGGTGGCGAT	CCCGGTGTCT
ACTGCGCCCG	ACGCCGTCCG	CACTTCCAGA	CGTACAGAG	GCCACCGCTA	GGGCCACAGA
4630	4640	4650	4660	4670	4680
TTGCCATGGC	GGCCGCCGTC	TGCGAGGCAG	TCGACAAAGGG	ACCGGCGGAA	TGGAAGTCGG
AACGGTACCG	CCGGCGGCAG	ACGCTCCCGT	AGCTGTTCCC	TGGCCGCCTT	ACCTTCAGCC
4690	4700	4710	4720	4730	4740
TTGAACTGGT	GATCACGCC	GGCGTGACCG	CGATGCTCGC	CGTTGCCGCC	CGCATCGCG
AACTTGACCA	CTAGTGCGGG	CCGCACCTGGC	GCTACGAGCG	GCAACGGCGG	CGTAGCCGC

FIGURE 8.5

18/151

4750	4760	4770	4780	4790	4800
CGCCGCTCGG	TCATGATTTC	TGTGCGATCT	CGCTTCCGA	CAATCTGAAG	CCCTGGGAAG
CGGGCGAGCC	AGTACTAAAG	ACACGCTAGA	GCGAAAAGGCT	GTAGACTTC	GGGACCCTTC
4810	4820	4830	4840	4850	4860
TCATCACCCG	CGGTCTCAGG	CTGGCGGCCG	AAGCGGGCTT	CGTCATTGCC	CTCTACAATC
AGTAGTGGGC	CGCAGAGTCC	GACCGCCGCG	TTCGCCGAA	GCAGTAACGG	GAGATGTTAG
4870	4880	4890	4900	4910	4920
CGATCAGCAA	GGCGCGGCC	TGGCAGCTCG	GTGAGGCCCTT	CGAGCTTCTG	CGCAGCGTTC
GCTAGTCGTT	CCCGCGCCGGG	ACCGTCGAGC	CACTCCGAA	GCTCGAAGAC	CGTTCGCAAG
4930	4940	4950	4960	4970	4980
TGCCGGCAAG	CGTTCCGGTC	ATCTTCGGCC	GTGCGGCCG	GCGGCCGGAC	GAACGGATCG
ACGGCCGTT	GCAAGGCCAG	TAGAACCGGG	CACGCCGGCC	CGCCGGCCTG	CTTGCCTAGC
4990	5000	5010	5020	5030	5040
CGGTGATGCC	GCTCGGCGAG	GCCGATGCCA	ACCGGCCGA	CATGGCGACC	TGCGTCATCA
GCCACTACGG	CGAGCCGCTC	CGGCTACGGT	TGGCGCCGCT	GTACCGCTGG	ACGCAGTAGT
5050	5060	5070	5080	5090	5100
TCGGCTCGCC	GGAGACGCGC	ATCGTCGAGC	GCGACGGCCA	ACCCGATCTC	GTCTACACAC
AGCCGAGCGG	CCTCTGCGCG	TAGCAGCTCG	CGCTGCCGGT	TGGGCTAGAG	CAGATGTGTG
5110	5120	5130	5140	5150	5160
CGCGCTCTA	TGCAGGGGCG	AGCCAGTGAG	CGATGCCGTT	GAGTGCCTCG	TCGCAACTGC
GCGCGAAGAT	ACGTCCCCGC	TCGGTCACTC	GCTACGCCA	CTCACGGAGC	AGCGTTGACG
5170	5180	5190	5200	5210	5220
CGACCGTCGG	CACGTCCGCG	GGCTTGCGCC	GCTCGACCAT	GATCACCTCG	ATGCCGAGCC
GCTGGCAGCC	GTGCAGGC	CCGAACGCGG	CGAGCTGGTA	CTAGTGGAGC	TACGGCTCGG
5230	5240	5250	5260	5270	5280
GGCGCGCTGC	GGCAATCTTG	CCGTAGGTGG	CGCTGCCACC	GCTGTTCTTG	GCGACGATCA
CCGCGCGACG	CCGTTAGAAC	GGCATCCACC	GCGACGGTGG	CGACAAGAAC	CGCTGCTAGT
5290	5300	5310	5320	5330	5340
CATCGATCTG	CCGACTCCTG	AGCAACGCGG	CTTCGTCGGC	TTCCGCAAAG	GGACCGGTG
GTAGCTAGAC	GGCTGAGGAC	TCGTTGCGCC	GAAGCAGCCG	AAGGCCTTTC	CCTGGCCAGC
5350	5360	5370	5380	5390	5400
CCAGGATCGC	CTCCCTGGTCG	GGCAGATTAA	GCGGCGCGT	CACCGGATCG	ACGCTGCCA
GGTCCTAGCG	GAGGACCAGC	CCGTCTAATT	CGCCGCCGCA	GTGGCCTAGC	TGCGACGCC
5410	5420	5430	5440	5450	5460
TGACGTAGCT	GTGCTGCCG	GCGACCTCGA	AGTGGAAAGC	TTCCCTGTCGA	CCTATGCCA
ACTGCATCGA	CACGACGCCG	CGCTGGAGCT	TCACCTTTCG	AAGGACAGCT	GGATAGCGGT
5470	5480	5490	5500	5510	5520
GGAAGACCGC	GCGTCGCCGA	TCACCGAGCG	CGCTGACGGC	CTCGACACG	CTATCGACAG
CCTTCTGCGC	CGCAGCGGCT	AGTGGCTCG	GCGACTGCCG	GAGCTGTTGC	GATAGCTGTC
5530	5540	5550	5560	5570	5580
CAGTCCAGCG	GTCGCCAGGC	AGGGGCACCC	ATTCCGGTCG	GCGGAGGGCG	ATAAGCGCAA
GTCAGGTGCG	CAGCGGTCCG	TCCCCGTGCG	TAAGGCCAGC	CGCCTCCCGC	TATTGCGGTT
5590	5600	5610	5620	5630	5640
CGCCGGTTCT	TTGCGCTGCG	TCCCGGGCGT	TGTGCGAAAT	GCGTGCGGCA	AAGGGGTGCG
GCGGCCAAGA	AACCGACGC	AGGCGCCGCA	ACACGTTTA	CGCACGCCGT	TTCCCCACGC
5650	5660	5670	5680	5690	5700
TCGCATCGAC	CAGCAGCGC	ATGTTTCTG	CATGCACGAA	ATGCGCCAGC	CCATCGGCC

19/151

AGCGTAGCTG GTCGTGCGC TACAAAAGCA GTACGTGCTT TACGCGGTGCG GGTAGCCGCG  
 5710 5720 5730 5740 5750 5760  
 CGCCAAAGCC GCCGATGCGC GTCTTGACCG GCTGGGCCG CGGGTCCGCG GTGCGGCCGG  
 GCGGTTTCGG CGGCTACGCG CAGAACTGGC CGACGCCGGC GCCCAGGCGC CACGCCGGCC  
 5770 5780 5790 5800 5810 5820  
 CCAGCGAGAT GGCGGTGTCG TAGCGGACAT CTTCGGCAA GCGGCGCGCG AGTTCCGCGTGC  
 GGTCGCTCTA CCGCCACAGC ATCGCCTGTA GAAGCCGGT CGCCGCGCGC TCAAGCGCAC  
 5830 5840 5850 5860 5870 5880  
 CCTCGGTGGT GCCACCCAGA ATCAGAACAT GAGGTTTTTC CATGGCTGAC GTGTGAACA  
 GGAGCCACCA CGGTGGGTCT TAGTCTTATG CTCCAAAAAG GTACCGACTG CACAGCTTGT  
 5890 5900 5910 5920 5930 5940  
 GCGAACCGC CATAGTCTCC CCCTGGCTGA CCGTCATCGG TATCGGTGAG GATGGTGTAG  
 CGCTTGGCG GTATCAGAGG GGGACCGACT GGCAGTAGCC ATAGCCACTC CTACCACATC  
 5950 5960 5970 5980 5990 6000  
 CGGGTCTCGG CGACGAGGCC AAGCGGCTGA TCGCCGAAGC GCGGCTGTC TACGGCGGCC  
 GCCCAGAGCC GCTGCTCCGG TTCGCGACT AGCGGCTTCG CGGCCAGCAG ATGCCGCCGG  
 6010 6020 6030 6040 6050 6060  
 ATCGTCATCT GGAGCTCGCC GCCTCCCTCA TCACCGGCCA AGCGCACAAT TGGCTAAGCC  
 TAGCAGTAGA CCTCGAGCGG CGGAGGGAGT AGTGGCCGCT TCGCGTGTAA ACCGATTGG  
 6070 6080 6090 6100 6110 6120  
 CCCTCGAACG CTCGGTCGTC GAGATCGTCG CGCGTCGCGG CAGCCCGGTG GTGGTGTGTT  
 GGGAGCTTGC GAGCCAGCAG CTCTAGCAGC GCGCAGCGCC GTCGGGCCAC CACCACGAAC  
 6130 6140 6150 6160 6170 6180  
 CCTCGGGCGA CCCGTTCTTC TTTCGGCGTCG GCGTACGCT GCGCAGCGCC ATCGCCCTCGG  
 GGAGCCCGCT GGGCAAGAAC AAGCCGCAGC CGCACTGCGA CCGCGCGGGCG TAGCGGAGCC  
 6190 6200 6210 6220 6230 6240  
 CCGAAATACG CACGCTTCGG GCGCCGTCG CGATCAGTCT TGCGCCTCG CGCCTCGGCT  
 GGCTTATGC GTGCGAAGGC CGCGGCAGCA GCTAGTCAGA ACGGCGGAGC GCGGAGCCGA  
 6250 6260 6270 6280 6290 6300  
 GGGCGCTGCA GGATGCGACG CTCGTCTCCG TACATGGCGC GCGCCTGGAT CTGGTGCAG  
 CCCGCGACGT CCTACGCTGC GAGCAGAGGC ATGTACCCGC CGCGCACCTA GACCACGCTG  
 6310 6320 6330 6340 6350 6360  
 CGCATTTGCA TCCGGGGCG CGTGTGCTTA CGCTCACGTC GGACGGTGC GGTCCGCGAG  
 GCGTAAACGT AGGCCCCCGC GCACACGAAT GCGAGTGCAG CCTGCCACGC CCAGGGCGCTC  
 6370 6380 6390 6400 6410 6420  
 ACCTTGCCGA GCTTCTGGTT TCAAGCGGCT TCGGTCACTC GCGACTGACC GTGCTCGAAG  
 TGGAACGGCT CGAACGACAA AGTCGCGCA AGCCAGTCAG CGCTGACTGG CACGAGCTTC  
 6430 6440 6450 6460 6470 6480  
 CGCTGGCGG CGCCGGCGAA CGGGTGACGA CGCAGATCGC CGCGCGCTTC ATGCTCGGCC  
 GCGACCCGCC CGGGCCGCTT GCCCACTGCT GCGTCTAGCG CGCGCGGAAG TACGAGCCGG  
 6490 6500 6510 6520 6530 6540  
 TCGTGCATCC TTTGAACGTC TGCGCCATTG AGGTGSCGGC CGACGAGGGC GCGCGCATCC  
 AGCACGTAGG AAACCTGCA ACACGGTAAC TCCACCGCCG GCTGCTCCCG CGCGCGTAGG  
 6550 6560 6570 6580 6590 6600  
 TGCCGCTTGC CGCCGGCCGC GACGATGCGC TGTCGAACA TGACGGGCAG ATCACCAAGC  
 ACGGCGAACG CGGGCCGGCG CTGCTACCGC ACAAGCTTGT ACTGCCCGTC TAGTGGTTCG  
 6610 6620 6630 6640 6650 6660

20/151

GCGAGGTGCG GGCCTGACG CTGTCGGCAC TCGCACCGCG CAAGGGCGAA CTGCTATGGG  
 CGCTCCACGC CCGCGACTGC GACAGCCGT AGCGTGGCGC GTTCCCCTT GACGATAACCC

6670 6680 6690 6700 6710 6720  
 ACATCGGGCG CGGCTCCGGC TCGATCGGC TCGAATGGAT GCTCGCCGAT CCGACCATGC  
 TGTAGCCGCC CGCGAGGCCG AGCTAGCCGT AGCTTACCTA CGAGCGGCTA GGCTGGTACG

6730 6740 6750 6760 6770 6780  
 AGGCGATCAC CATCGAGGTT GAGCCGGAGC GGGCAGCGCG CATCGGCCGC AACGCGACGA  
 TCCGCTAGTG GTAGCTCAA CTCGGCCTCG CCCGTCGCGC GTAGCCGGCG TTGCGCTGCT

6790 6800 6810 6820 6830 6840  
 TGTTCGGGCGT GCCCCGGCTG ACGGTTGTCG AAGGCGAGGC GCCGGCGGCG CTTGCCGGCC  
 ACAAGCCGCA CGGGCCCGAC TGCCAACAGC TTCCGCTCCG CGGCCGCGC GAACGGCCGG

6850 6860 6870 6880 6890 6900  
 TGCCACAAAC GGACCGCAGTC TTCATCGGC GCGGCGGCAG CGAACAGCGC GTCATGGAAG  
 ACGGTGTGG CCTGCCTAG AAGTAGCCGC CGCCGCCGTC GCTTCTGCCG CAGTACCTTC

6910 6920 6930 6940 6950 6960  
 CAGCGATCGA GGCCTCAAG TCAGGCGGAC GGCTGGTTGC CAACCGGGTG ACGACGGACA  
 GTCGCTAGCT CCGCGAGTT AGTCCGCCGT CGGACCAACG GTTGCGCCAC TGCTGCCTGT

6970 6980 6990 7000 7010 7020  
 TGGAAAGCGGT GCTGCTCGAT CATCACGCCG GGCTGGCGG TTCGCTGATC CGCATCGATA  
 ACCTTCGCCA CGACGAGCTA GTAGTGCCTG CCGAGCCGCC AAGCGACTAG GCGTAGCTAT

7030 7040 7050 7060 7070 7080  
 TCGCGCGTGC AGGACCCATC GGCAGCATGA CCGGCTGGAA GCCGGCCATG CCGGTACCC  
 AGCGCGCACG TCCTGGGTAG CCGCCGTACT GGCCGACCTT CGGCCGGTAC GGCCAGTGGG

7090 7100 7110 7120 7130 7140  
 AATGGTCGTG GACGAAGGGC TAAAGCAGTT CCAGCGAAAG TGTGACGCGG TTTTGCCTCC  
 TTACCAAGCAC CTGCTTCCCG ATTTCGTCAA GGTCGCTTC ACACTGCGCC AAAACGCAAG

7150 7160 7170 7180 7190 7200  
 GGAACCTGCGC AAGAAAAAGA AAGAGTAACC TATGACGGTA CATTTCATCG GCGCCGGCCC  
 CCTTGACGCCG TTCTTTTCT TTCTCAITGG ATACTGCCAT GTAAAGTAGC CGCGGCCGGG

7210 7220 7230 7240 7250 7260  
 GGGAGCCGCA GACCTGATCA CGGTGCGTGG TCGCGACCTG ATCGGGCGCT GCCCGGTCTG  
 CCCTCGGGGT CTGGACTAGT GCCACGCCACC AGCGCTGGAC TAGCCCGCGA CGGGCCAGAC

7270 7280 7290 7300 7310 7320  
 CCTTTACGCC GGCTCGATCG TCTCGCCGGA GCTGCTGCGA TATTGCCCCG CGGGCGCCCG  
 GGAAATGCCG CCGAGCTAGC AGAGCGGCCCT CGACGACGCT ATAACGGGGCG GCCCCGCCGGC

7330 7340 7350 7360 7370 7380  
 CATTGTGAT ACGGCGCCGA TGTCCCTCGA CGAGATCGAG GCGGAGTATG TGAAGGCCGA  
 GTAACAGCTA TGCCCGGGCT ACAGGGAGCT GCTCTAGCTC CGCCTCATAC ACTTCCGGCT

7390 7400 7410 7420 7430 7440  
 AGCCGAAGGG CTCGACGTGG CGCGGCTTCA TTGCGGCCGAC CTTTCGGTCT GGAGTGTGT  
 TCGGCTTCCC GAGCTGCACC GCGCCGAAGT AAGCCCGCTG GAAAGCCAGA CCTCACGACA

7450 7460 7470 7480 7490 7500  
 GGCGCAACAG ATCCGCCGGC TCGAGAAGCA TGGCATCGCC TATACGATGA CGCCGGCGT  
 CGGGCTTGTGTC TAGGCCGGCG AGCTCTTCGT ACCGTAGCGG ATATGCTACT CGGGCCCGCA  
 AGGAAGGAAA CGCCGCCCGCC GAAGTCGCGA GCCAGCGCTT AACTGGTAAG GCGGGCACCG

21/151

7570	7580	7590	7600	7610	7620
CCAGAGCCG	GTGCTGACCC	GCGTTTCGGG	CCGCCTCG	CCGATGCCGA	ACTCAGAAAC
GGTCTCGGAC	CACGACTGGG	CGCAAAGCCC	GGCGCGGAGC	GGCTACGGCT	TGAGTCTTG
7630	7640	7650	7660	7670	7680
GCTTTCCGCT	TTCGGGCGCTA	CGGGATCGAC	GCTGGCAATC	CACCTTGCAG	TCCATGCGCT
CGAAAAGCGA	AAGCCGCGAT	GCCTAGCTG	CGACCGTTAG	GTGGAACGCT	AGGTACCGCA
7690	7700	7710	7720	7730	7740
TCAGCAGGTG	GTCGAGGAAC	TGACGCCGCT	CTACGGTGCC	GAUTGCCCGG	TCGCCATCGT
AGTCGTCCAC	CAGCTCCTTG	ACTGCGGCGA	GATGCCACGG	CTGACGGGCC	AGCGGTAGCA
7750	7760	7770	7780	7790	7800
CGTCAAGGCC	TCCTGCCCGG	ACGAACCGCT	GGTGCCTGGC	ACGCTCGGTG	ACATGCCCGC
CGAGTTCCGG	AGGACCGGCC	TGCTTGCAGCA	CCACGCCCG	TGCGAGGCCAC	TGTAGCGGCC
7810	7820	7830	7840	7850	7860
CAAGGTGGCG	GAAGAGCCGA	TCGAGGCCAC	GGCGCTGATC	TTCGTCCGGTC	CGGGGCTCGA
GTTCCACCGC	CTTCTCGGCT	AGCTCGCGTG	CCGCGACTAG	AAGCAGCCAG	GCCCCGAGCT
7870	7880	7890	7900	7910	7920
AGCCCTCCGAT	TTCCGTAAA	GCTCGCTCTA	CGATCCCGCC	TATCAGCGGC	GCTTCAGAGG
TCGGAGGCTA	AAGGCACCTT	CGAGCGAGAT	GCTAGGGCGG	ATAGTCGCCG	CGAAGTCTCC
7930	7940	7950	7960	7970	7980
CGCGGGCGAA	TAGGCCGCAC	TCCCTCGGGG	GTCCGCTAA	GTTCCTCGCT	GAGAGGGTTT
CGCGCCGCTT	ATCCGGCGTG	AGGGAGCCCC	CAGCCGGATT	CAAAGGGCGA	CTCTCCAAA
7990	8000	8010	8020	8030	8040
TGAAACCTAT	TCTGCCGGTT	CTTCGCGCGG	CGGCCGCTGC	TTGAGCGGGG	CGCCGCGCTT
ACTTTGGATA	AGACGGCCAA	GAAGCGCGCC	GCCGGCGACG	AACTCGCCCT	CGGGCGCGAA
8050	8060	8070	8080	8090	8100
TTCCTCGACG	CGGTCCGGGT	AGAGCGCTGC	CTGTCCAAGC	AGCATCAGCG	TCACCGGGCGT
AAGGAGCTGC	GCCAGCGCCA	TCTCGCGACG	GACAGGTTCG	TCGTAGTCGC	AGTGGCCGCA
8110	8120	8130	8140	8150	8160
GGTGGCGACG	ACGAAGACGA	TGATCAGGAT	TTCGTGGAAT	ACCCAGCGGC	TCTGCAGCAC
CCACCGCTGC	TGCTTCTGCT	ACTAGTCCTA	AAGCACCTA	TGGGTCGCCG	AGACGTCGTG
8170	8180	8190	8200	8210	8220
GGCAAAGCGAG	ATGATAGAGG	CGGCGCAGAT	CATCAGTACG	CCGCCGCTGG	TCGCCAGCGT
CCGTTTCGTC	TACTATCTCC	GCCGCGTCTA	GTAGTCATGC	GGCGCGACCC	AGCGGTCGCA
8230	8240	8250	8260	8270	8280
CGGTGCGTGC	AGGCGCTCGT	AGAACGCTGGT	GAACCGGAGC	AAGCCGACGG	AGCCGATCAG
GCCACCGCACG	TCCGCGAGCA	TCTTCGACCA	CTTGGCCTCG	TTCCGCTGCC	TCGGCTAGTC
8290	8300	8310	8320	8330	8340
CGCCACTGCG	GCGCCGAGGA	CGGTGAGGCC	GCAGACGAGA	ACGGCTGCC	AGACGGGAAG
GCGGTGACGC	CGCGGCTCCT	GCCACTCGGG	CGTCTGCTCT	TGCGACGGG	TCTGCCCTC
8350	8360	8370	8380	8390	8400
GTCGGTGAGG	TGGCTCATTC	GATGATCTCC	CCGCGCATCA	GGAACTTGCC	GAAGGGCGATC
CAGCCACTCC	ACCGAGTAAG	CTACTAGAGG	GGCGCGTAGT	CCTTGAACGG	CTTCCGCTAG
8410	8420	8430	8440	8450	8460
GACGAGACGA	AGCCGATCAA	AGCCACGATC	AGGGCGGACT	CGAAATAGAG	CGAGTTGGCC
CTGCTCTGCT	TCGGCTAGTT	TCGGTGCTAG	TCCCGCCTGA	GCTTTATCTC	GCTCAACCGG
8470	8480	8490	8500	8510	8520
GTGCGGATGC	CGAAGGTCAA	GAGCATCAGC	ATGGCGTTGA	TATAGAGCGT	GTCGAGGCCG
CACGCCTACG	GCTTCCAGTT	CTCGTAGTCG	TACCGCAACT	ATATCTCGCA	CAGCTCCGGC

FIGURE 8.9

22/151

8530        8540        8550        8560        8570        8580  
AGGATACGGT CCTGGCGCG CGGTCCCCTC ACCATGCCAT AGAAGGCAAA AGCCATCGCC  
TCCTATGCCA GGACCCGCGC GCCAGGGGAG TGGTAGGCTA TCTTCCGTTT TCGGTAGCGG

8590        8600        8610        8620        8630        8640  
AGGCCGAGCA TGATCTGGGC AATCAGGATC GACCAGATGA TTGAAAGTTC CATCATCCGA  
TCCGGCTCGT ACTAGACCCG TTAGTCCTAG CTGGTCTACT AACTTTCAAG GTAGTAGGCT

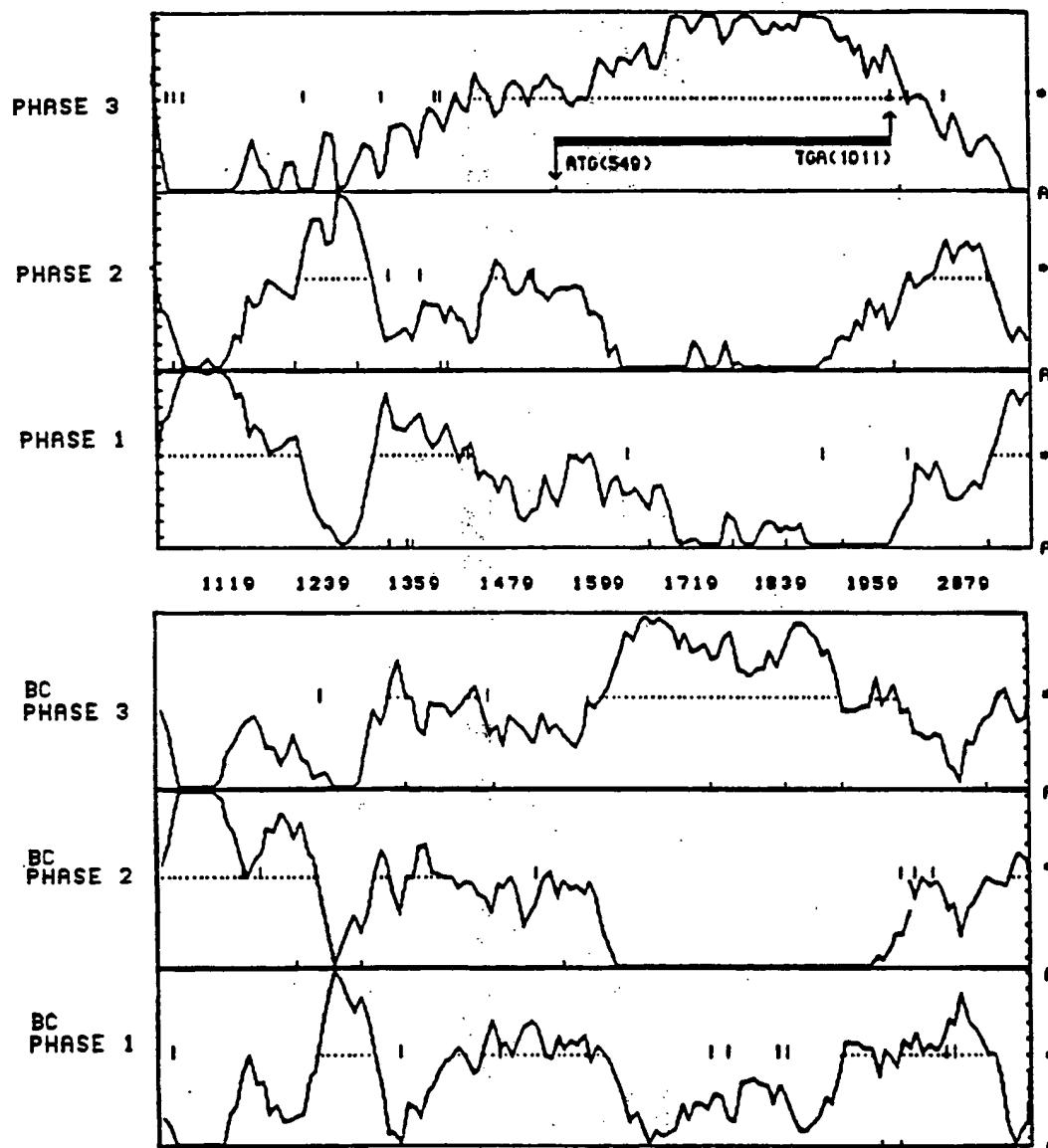
8650        8660        8670        8680        8690        8700  
ATATCTCCTT CAGGGCGGTC TCATAGCGCT TGACCGTATC GAGCCAGATG TCCTCGTTCT  
TATAGAGGAA GTCCCGCCAG AGTATCGCGA ACTGGCATAG CTCGGTCTAC AGGAGCAAGA

8710        8720        8730        8740        8750        8760  
CCATGTCGAG CACGTGGAAG AGCAGGGACT TGCAGCCGCG ATCCGGGGAA TTC  
GGTACAGCTC GTGCACCTTC TCGTCCCTGA ACGCCGGCGC TAGGCCCCTT AAG

FIGURE 8.10

23 / 151

SEQUENCE : FRAGMENT DE 5398 BP DE 1 A 1200 LONGUEUR = 1200



▲=ATG  
\* STOP  
BC BRIN COMPLEMENTAIRE

PHASE OUVERTE !

FIGURE 9.1

24/151

SEQUENCE: FRAGMENT DE 5398 BP DE 1000 A 2200 LONGUEUR = 1201 BP

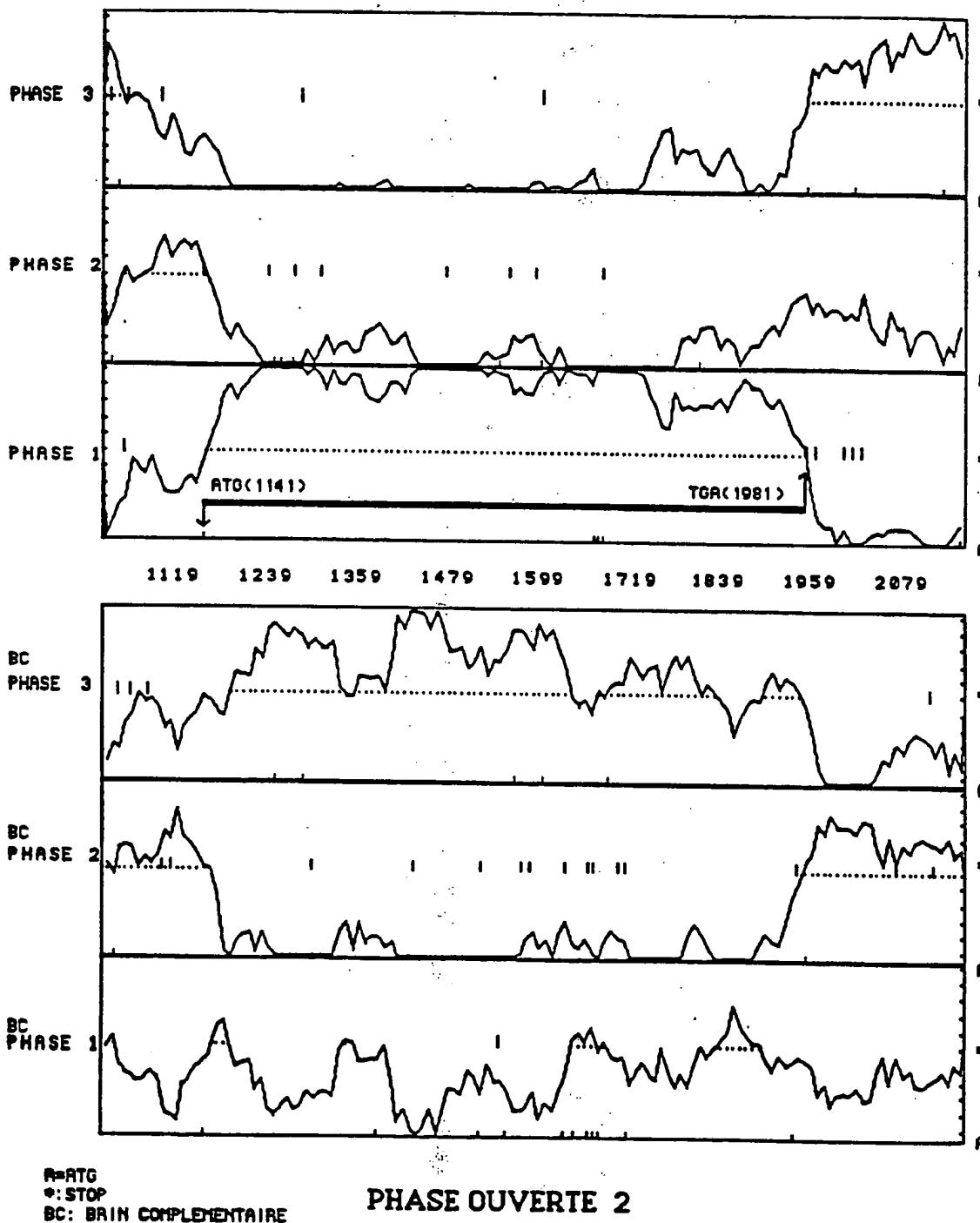
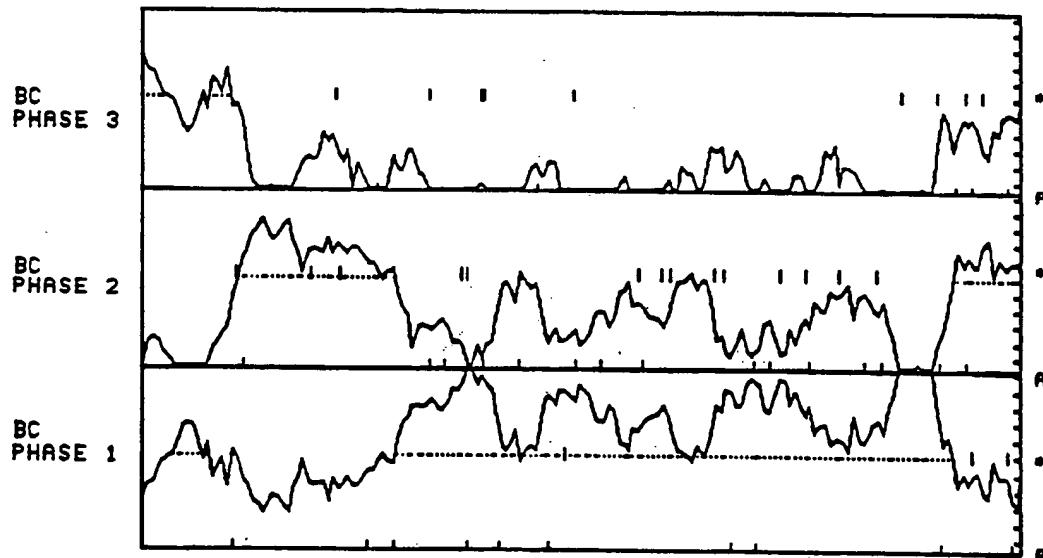
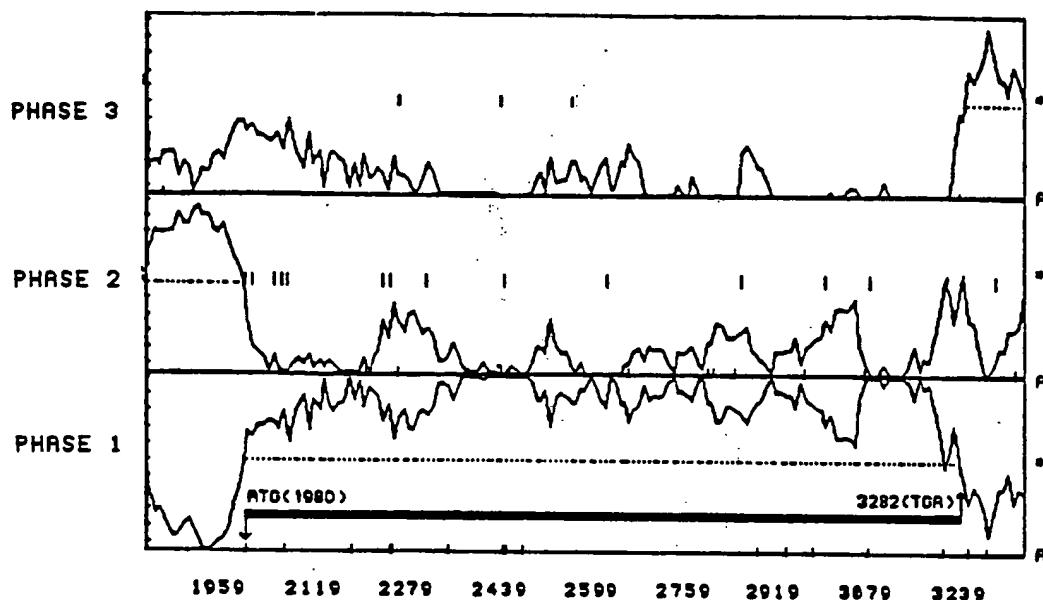
**PHASE OUVERTE 2**

FIGURE 9.2

**FEUILLE DE REMPLACEMENT**

25 / 151

SEQUENCE : FRAGMENT DE 5398 BP DE 1800 A 3400 LONGUEUR = 1601



R=ATG  
\*=STOP

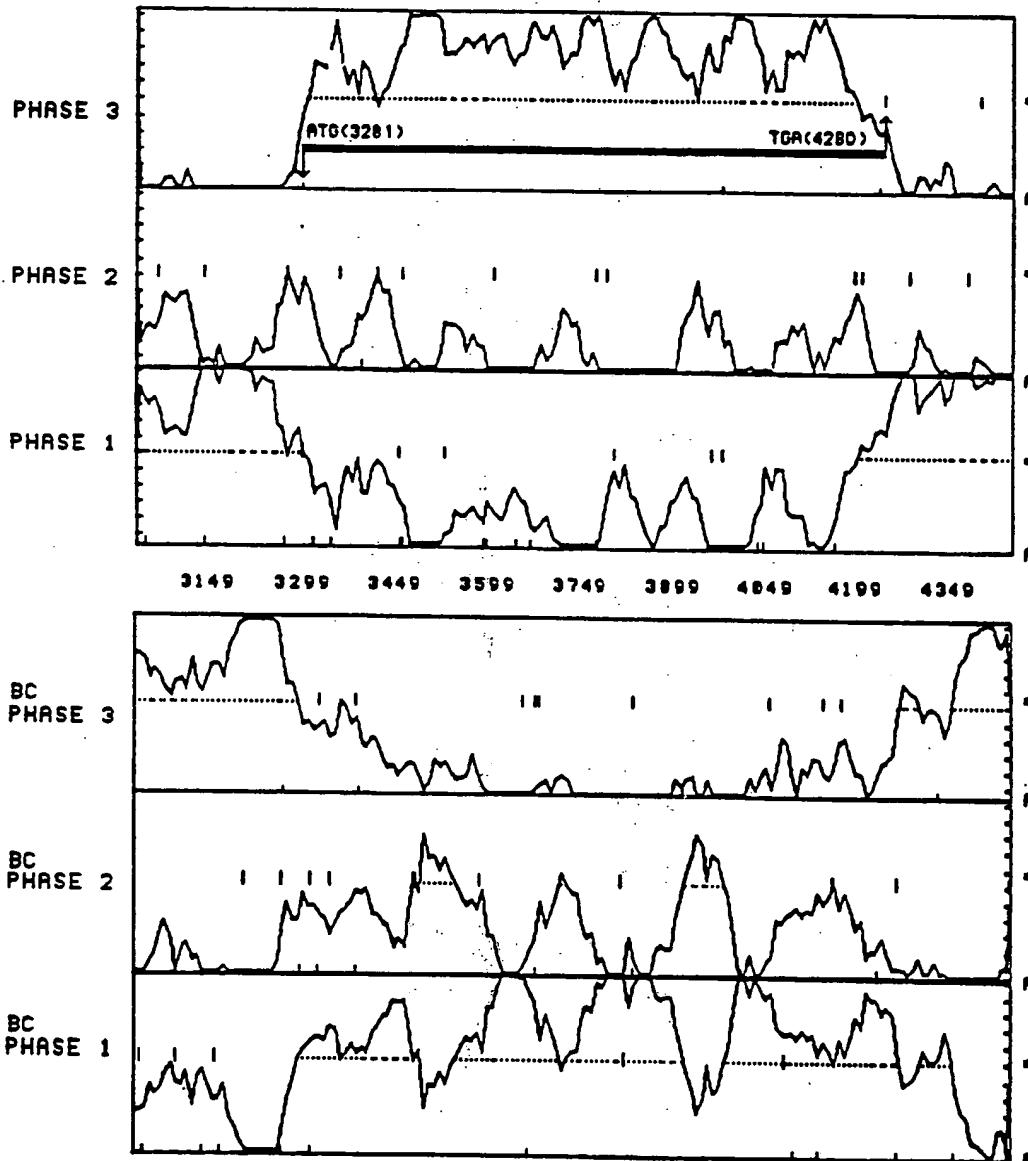
BC BRIN COMPLEMENTAIRE

## PHASE OUVERTE 3

FIGURE 9.3

26 / 151

SEQUENCE : FRAGMENT DE 5398 BP DE 3000 A 4500 LONGUEUR = 1501



## PHASE OUVERTE 4

FIGURE 9.4

FEUILLE DE REMplacement

27 / 151

SEQUENCE : FRAGMENT DE 5398 BP DE 3800 A 5398 LONGUEUR = 1599

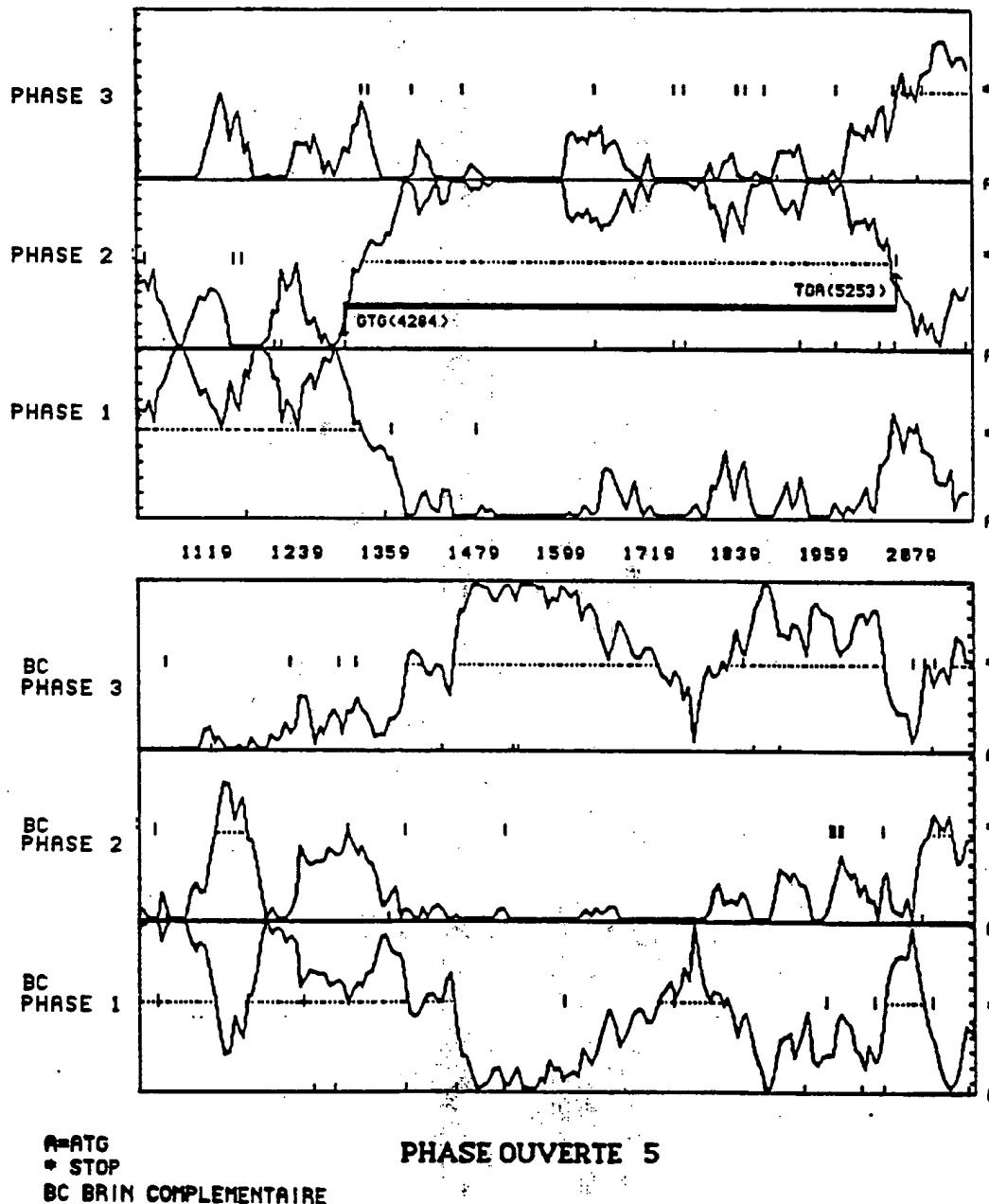
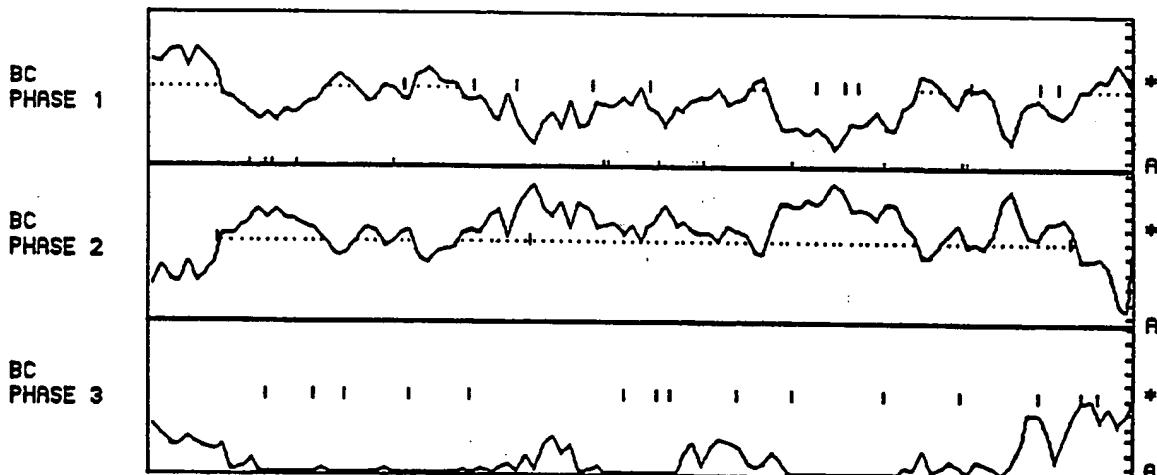
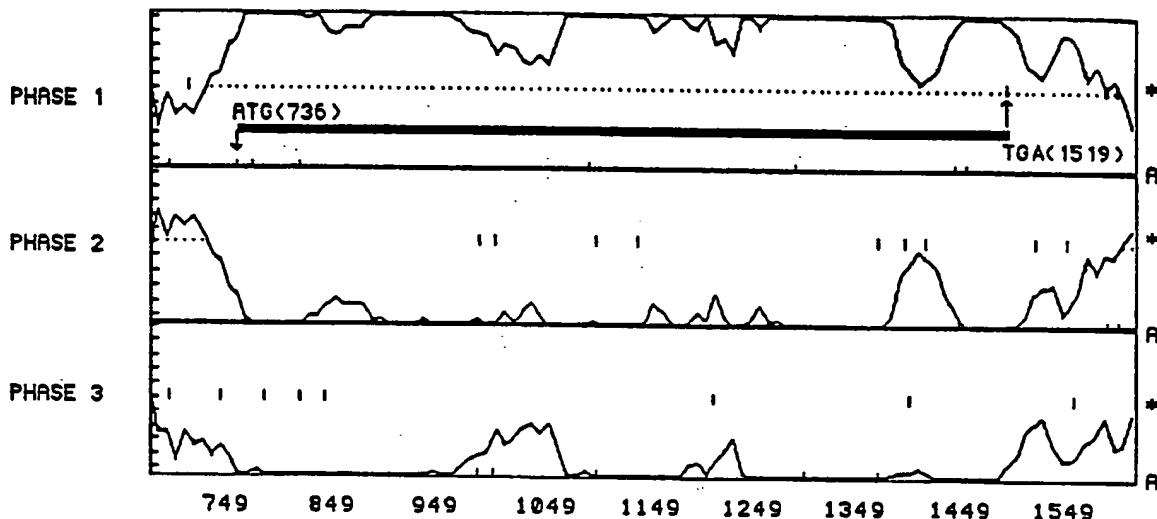


FIGURE 9.5

## **FEUILLE DE REMPLACEMENT**

28 / 151

SEQUENCE : FRAGMENT DE 8753 BP DE 650 A 1650 LONGUEUR = 1001



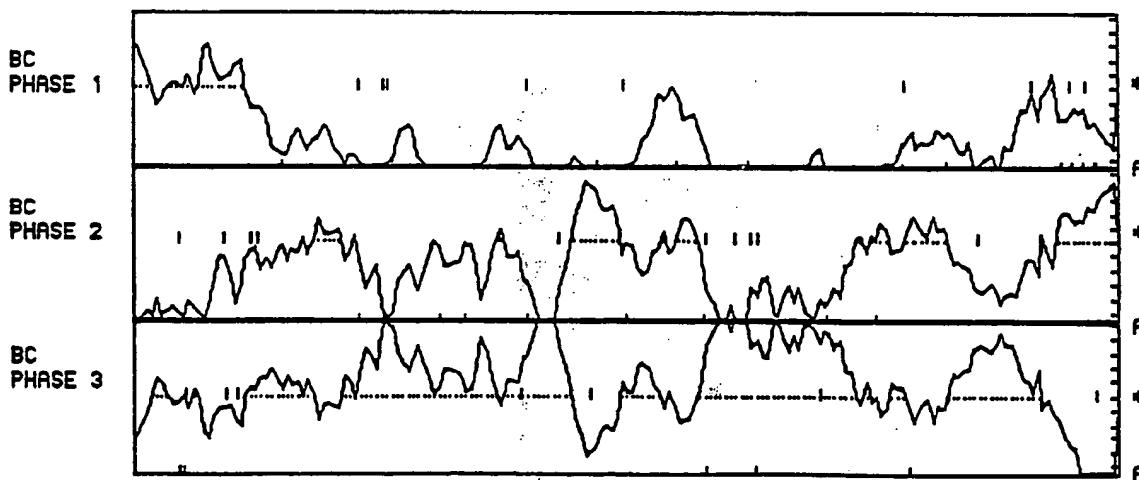
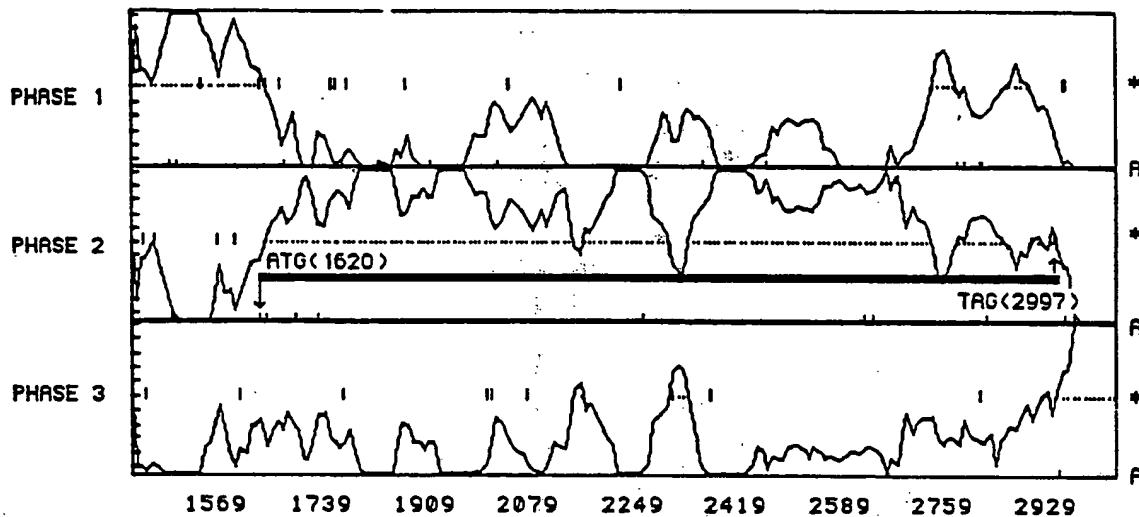
## PHASE OUVERTE 6

A=ATG  
\*=STOP  
BC BRIN COMPLEMENTAIRE

FIGURE 10.1

29 / 15 /

SEQUENCE : FRAGMENT DE 8753 BP DE 1400 A 3100 LONGUEUR = 1701



A=ATG  
\* STOP  
BRIN COMPLEMENTAIRE

PHASE OUVERTE 7

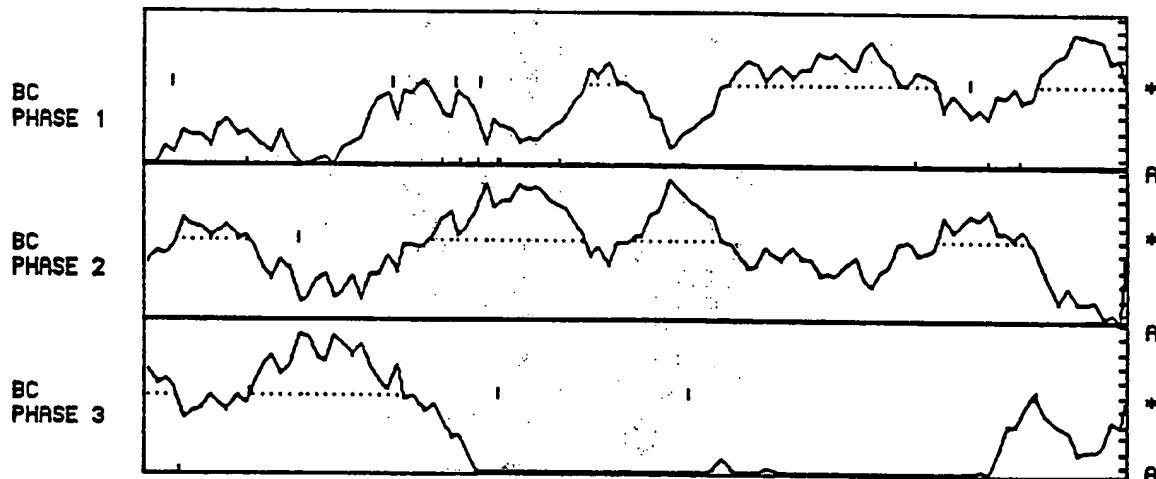
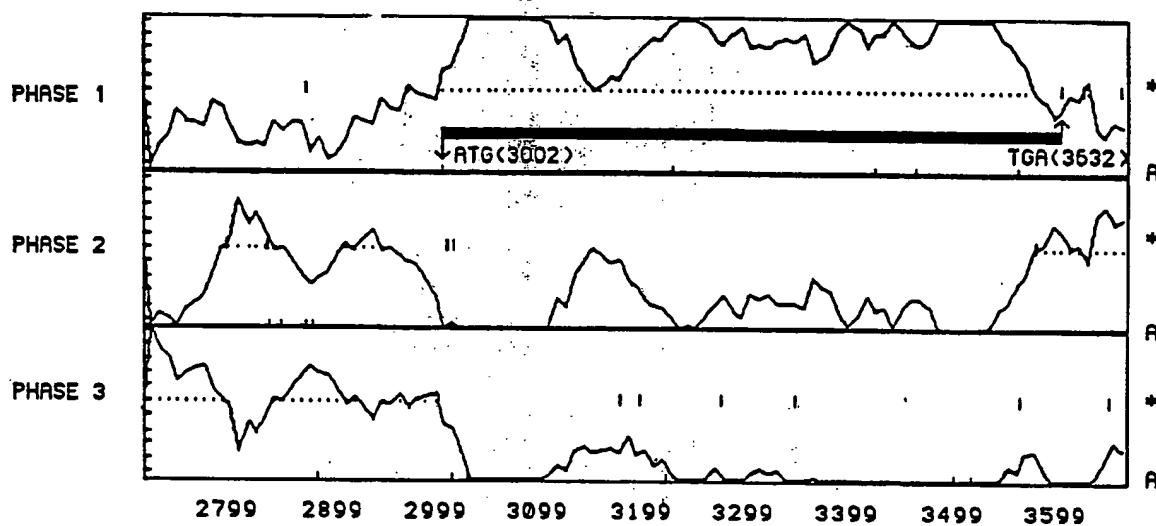
FIGURE 10.2

FEUILLE DE REMPLACEMENT

PL. 30 / 15 /

SEQUENCE : FRAGMENT DE 8753 BP DE 2700 A 3700

LONGUEUR = 1001



## PHASE OUVERTE 8

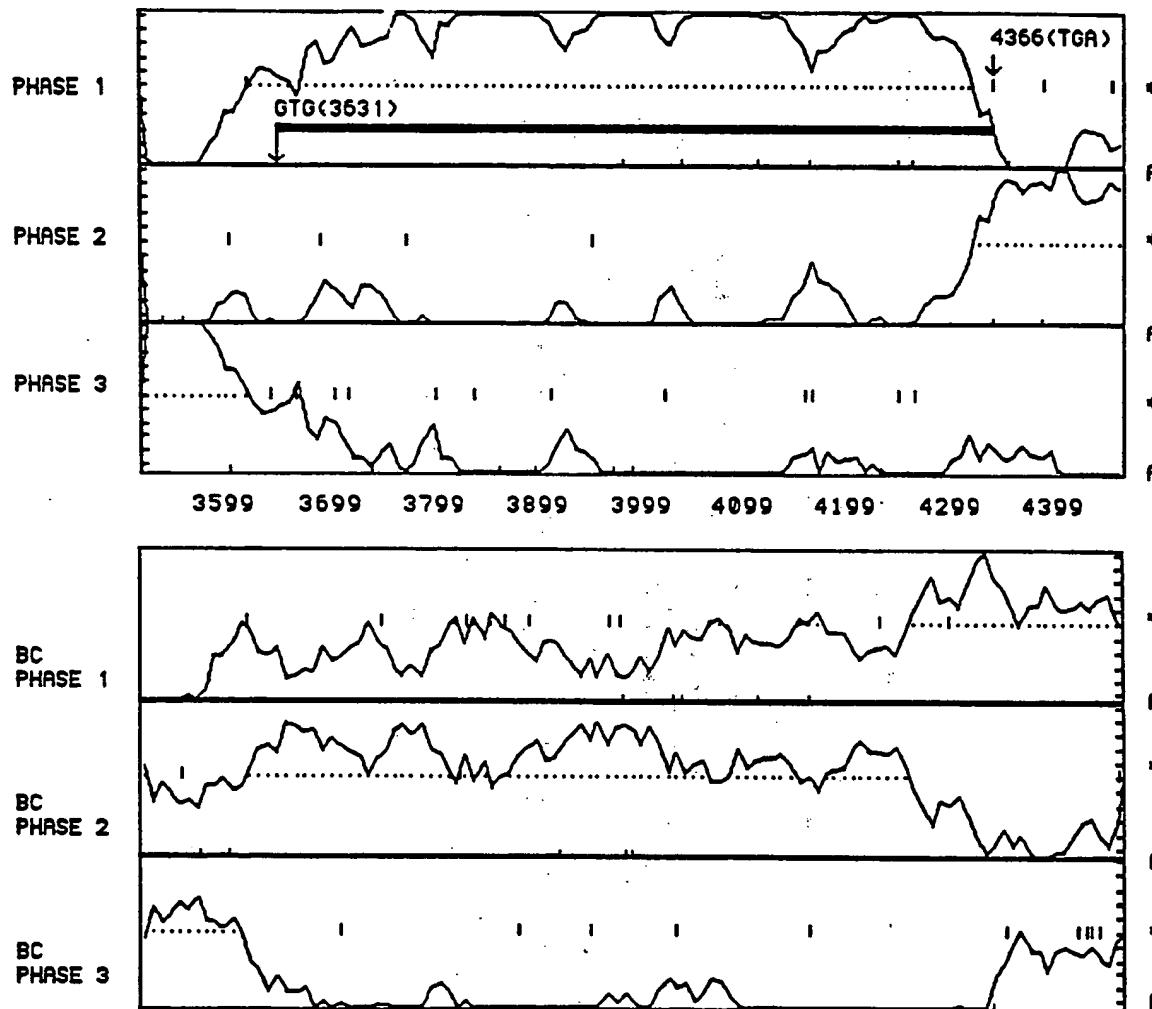
A=ATG  
\*= STOP  
BC BRIN COMPLEMENTAIRE

FIGURE 10.3

FEUILLE DE REMplacement

PL. 31/151

SEQUENCE : FRAGMENT DE 8753 BP DE 3500 A 4500 LONGUEUR = 1001



## PHASE OUVERTE 9

FIGURE 10.4

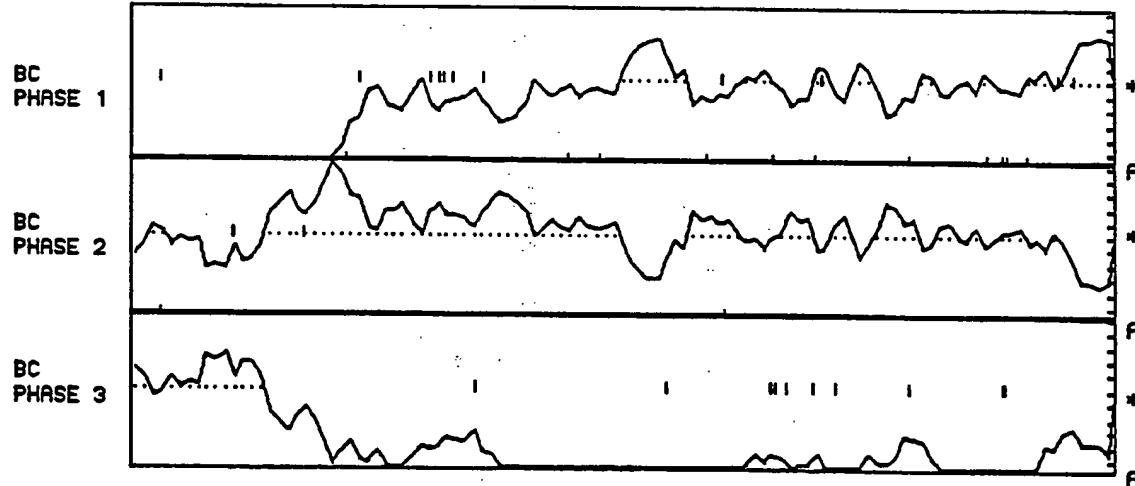
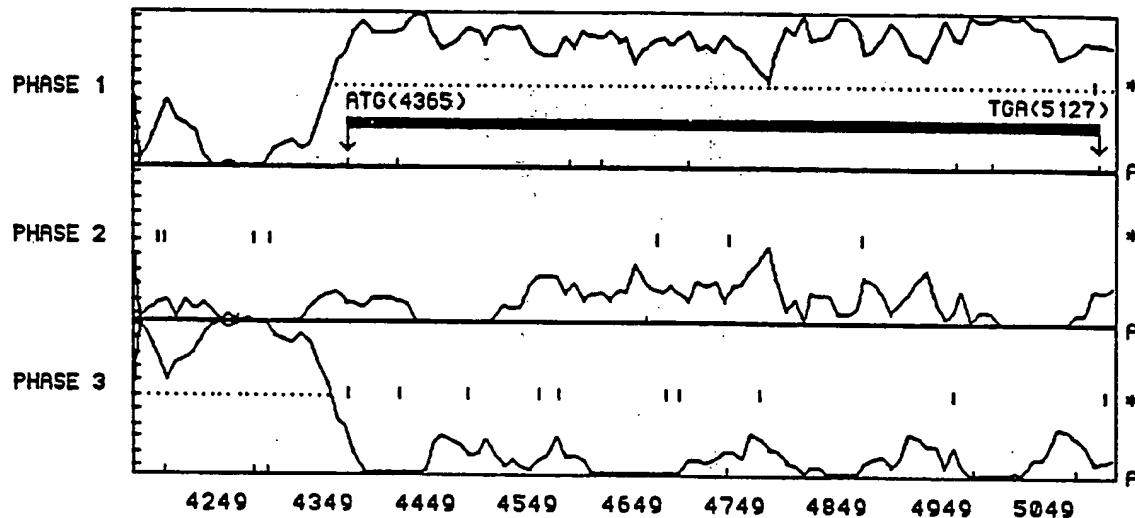
A=ATG  
\* STOP  
BC BRIN COMPLEMENTAIRE

FIGURE 10.4

FEUILLE DE REMplacement

PL. 32 / 151

SEQUENCE : FRAGMENT DE 8753 BP DE 4150 A 5150 LONGUEUR = 1001



## PHASE OUVERTE 10

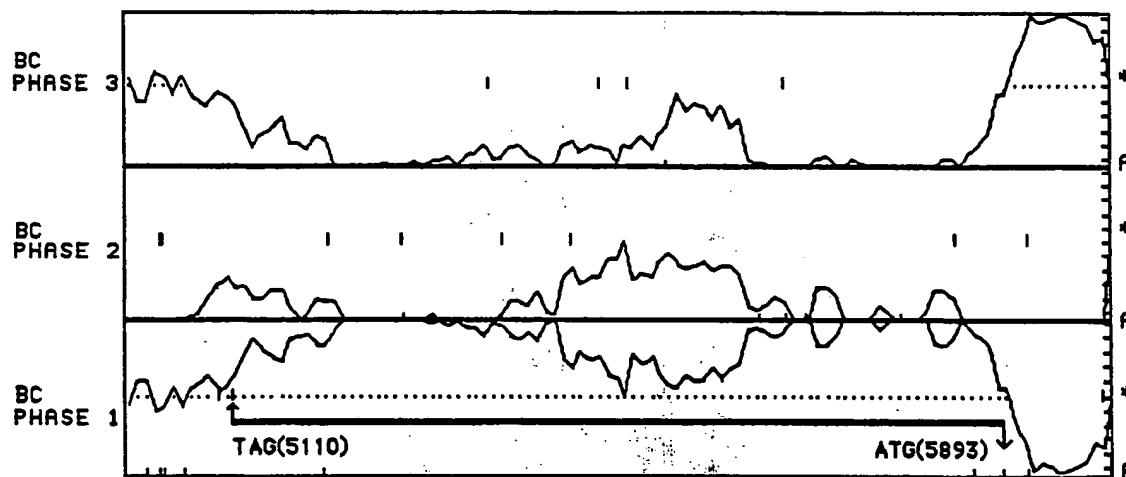
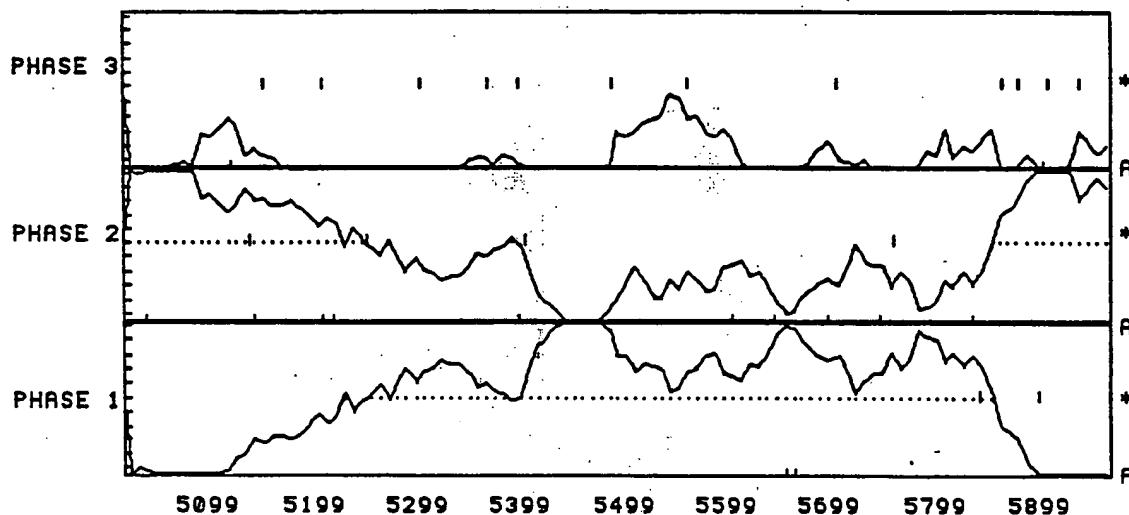
A=ATG  
\*= STOP  
BC BRIN COMPLEMENTAIRE

FIGURE 10.5

PL. 33 / 151

SEQUENCE : 253TOT

DE 5000 A 6000 LONGUEUR = 1001



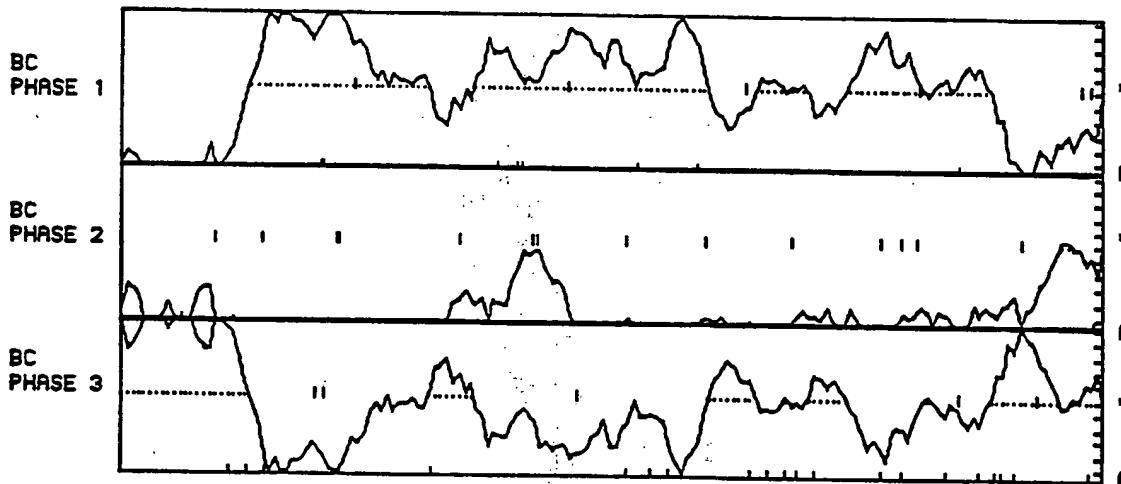
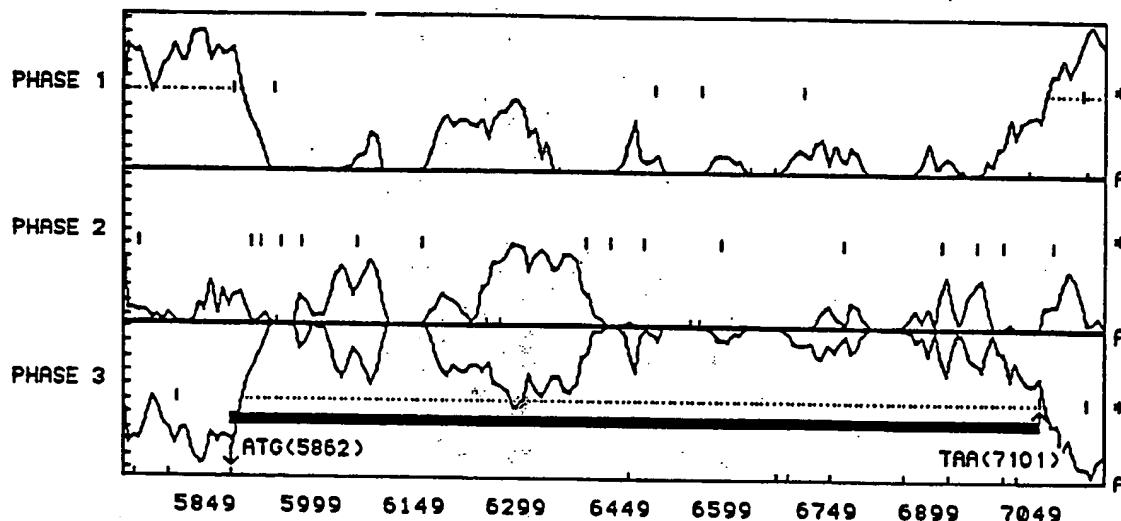
PHASE OUVERTE 11

FIGURE 10.6

FEUILLE DE REMPLACEMENT

PL. 34 / 151

SEQUENCE : FRAGMENT DE 8753 BP DE 5700 A 7200 LONGUEUR = 1501



## PHASE OUVERTE 12

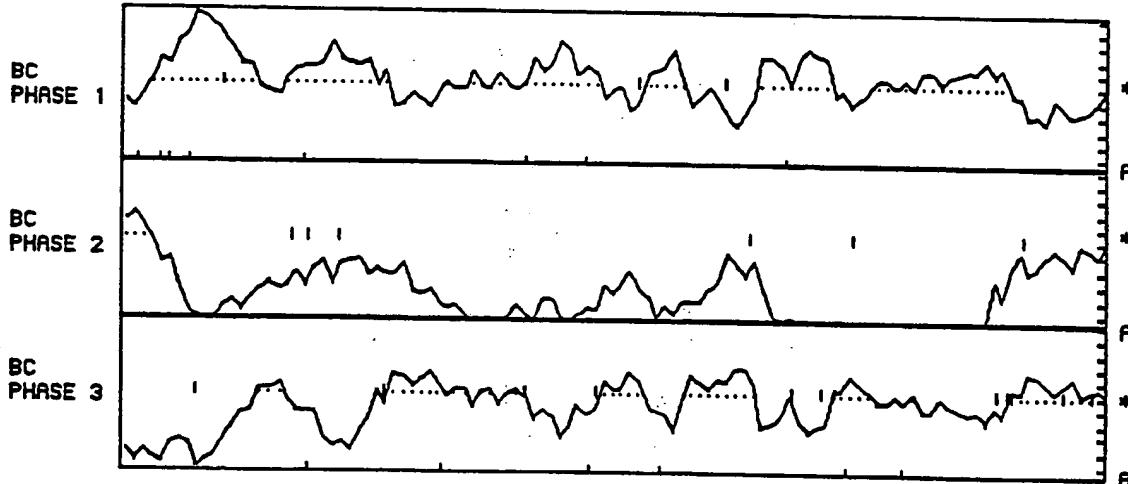
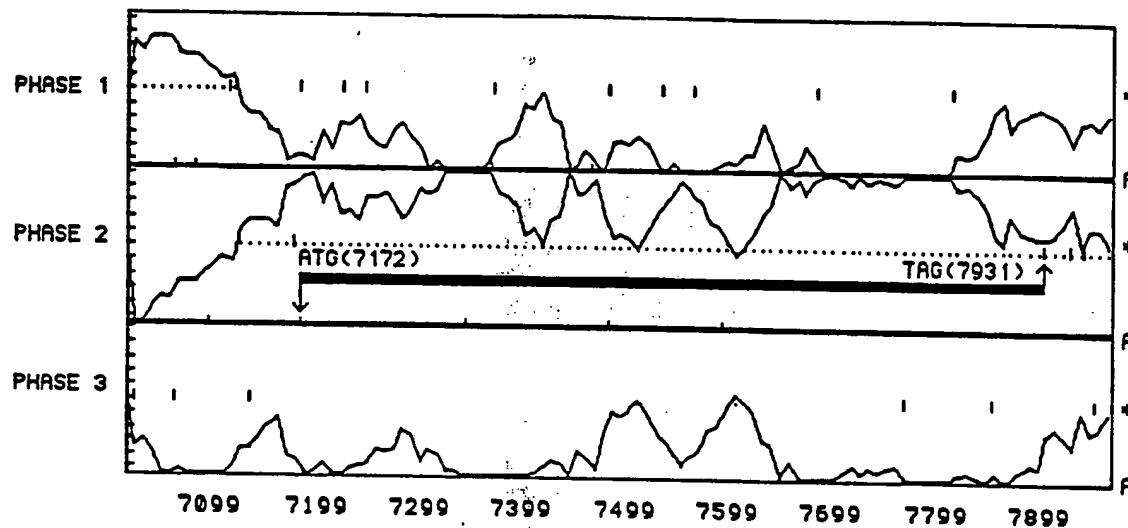
A=ATG  
\* STOP  
BC BRIN COMPLEMENTAIRE

FIGURE 10.7

FEUILLE DE REMPLACEMENT

PL. 35 / 151.

SEQUENCE : FRAGMENT DE 8753 BP DE 7000 A 8000 LONGUEUR = 1001



## PHASE OUVERTE 13

A=ATG  
\* STOP  
BC BRIN COMPLEMENTAIRE

FIGURE 10.8

36 / 151

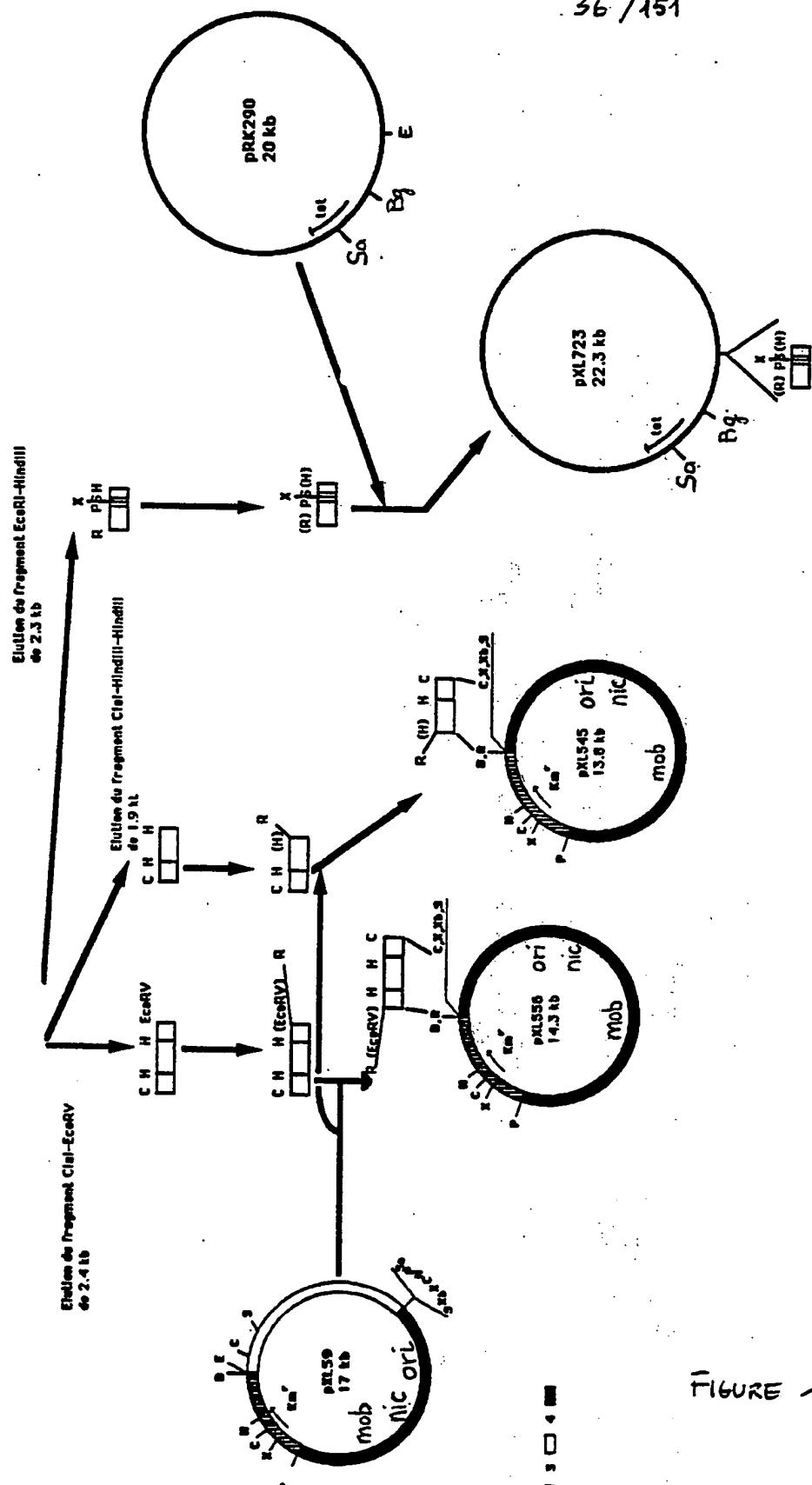


FIGURE 11

37 / A51

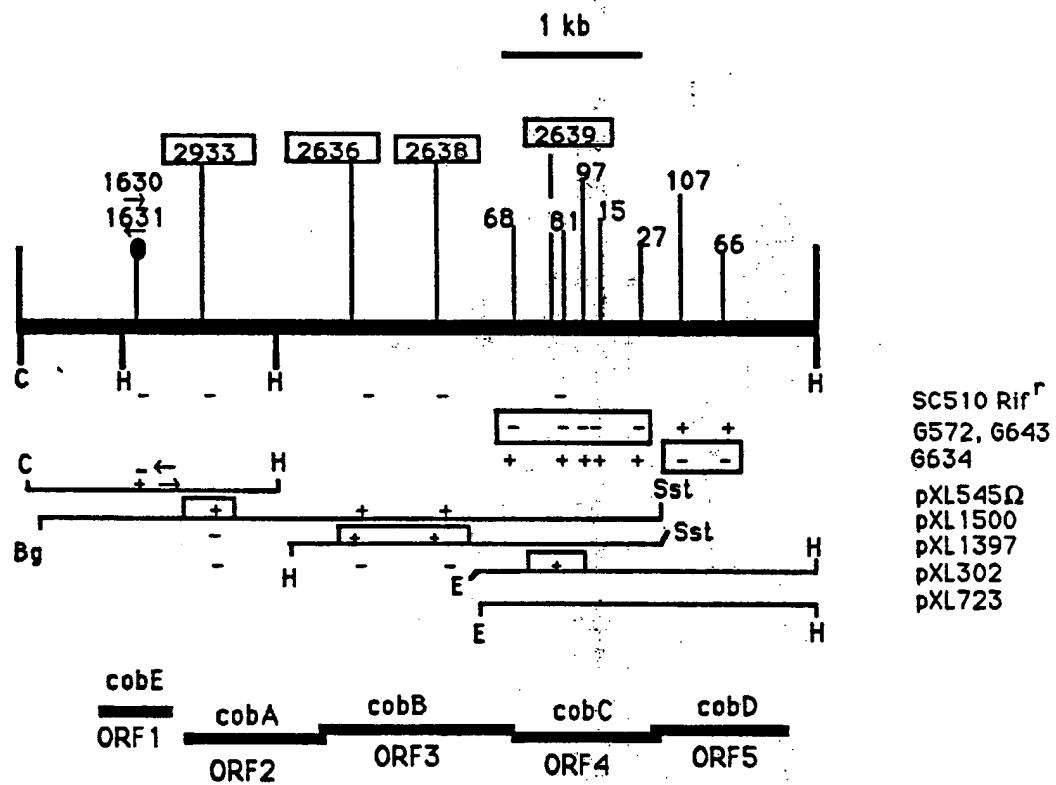


FIGURE 12

38 / 151

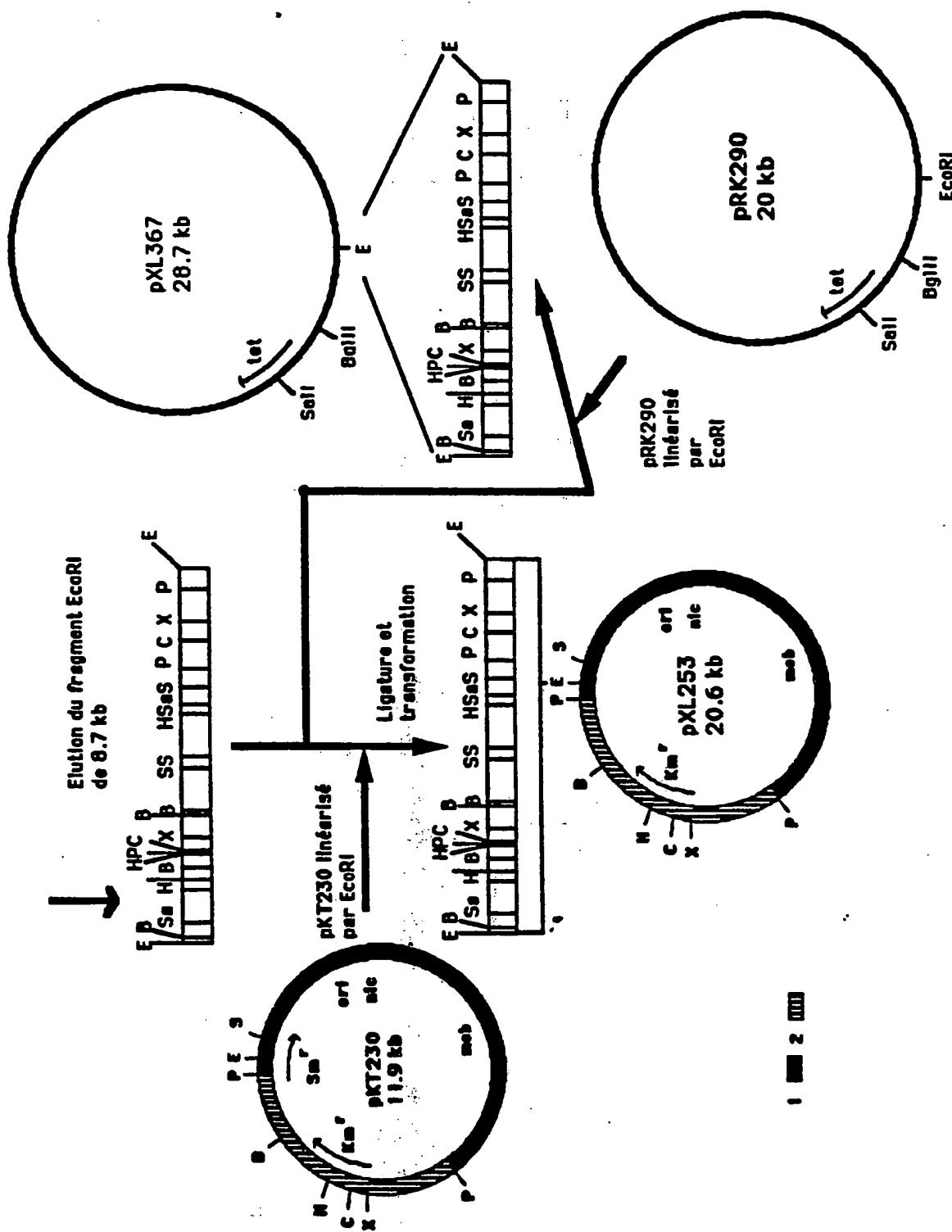


FIGURE 13

39 / 151

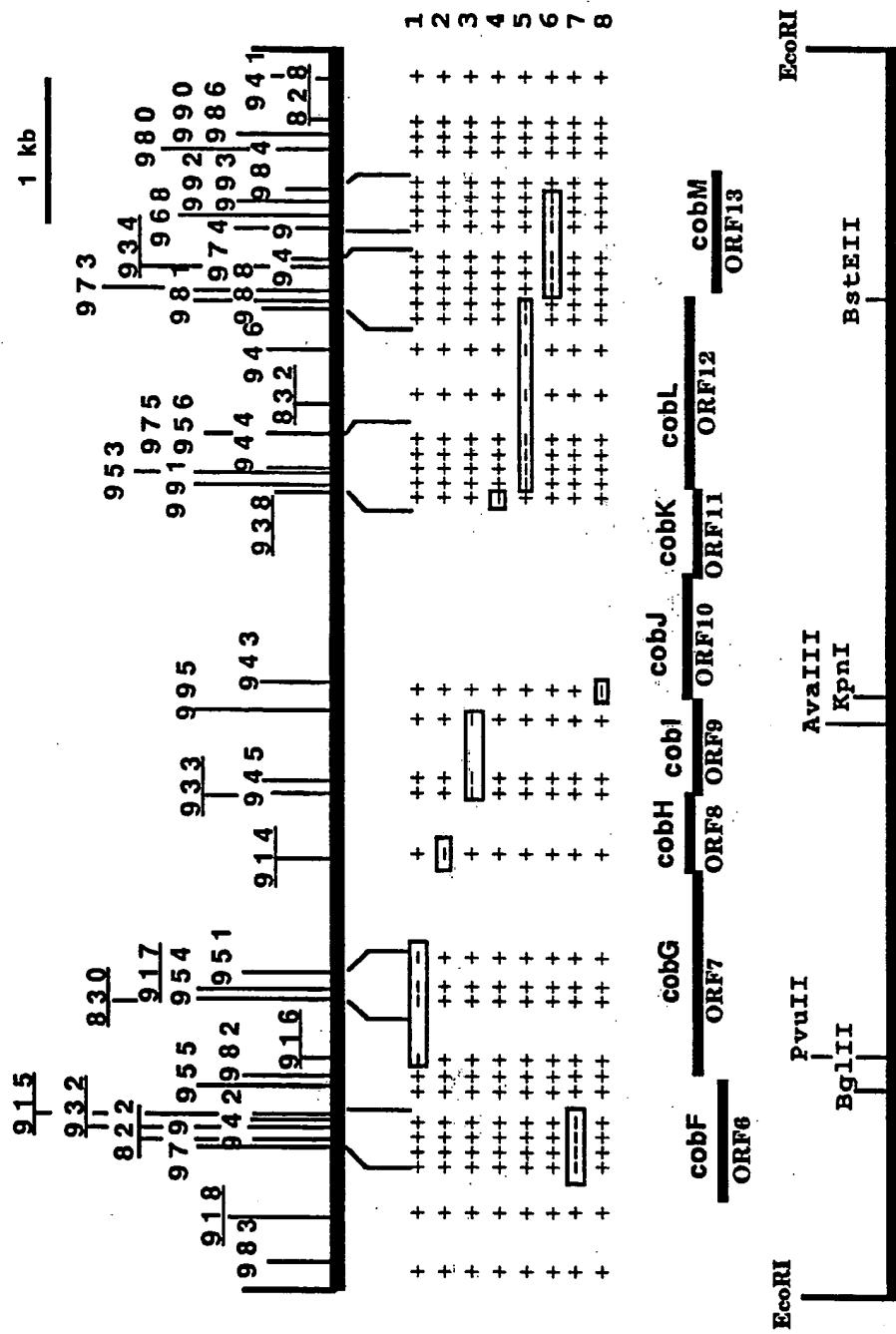


FIGURE 14

40 / 151

GENE cobA ET PROTEINE COBASEQUENCE DU FRAGMENT ClaI-HindIII-HindIII DE 5396 BP, DE 1141 À 1980

Met Ile Asp Asp Leu Phe Ala Gly Leu Pro Ala Leu Glu Lys Gly Ser Val Trp Leu Val Gly Ala Gly Pro Gly  
 ATGATCGACGACTCTTGGCGATTGCCGGCTCGAAAAAGGTTGGCTGGCTGGCTGGCGCCGGCCCCGGC  
 1141 1151 1161 1171 1181 1191 1201  
 Asp Pro Gly Leu Leu Thr Leu His Ala Ala Asn Ala Leu Arg Gln Ala Asp Val Ile Val His Asp Ala Leu Val  
 GATCCCGGCCTGTTGACGCTGCATGCGCCAATGCGCTGCGCCAGGC GGATGTGATCGTCATGATGCGCTGGTC  
 1216 1226 1236 1246 1256 1266 1276  
 Asn Glu Asp Cys Leu Lys Leu Ala Arg Pro Gly Ala Val Leu Glu Phe Ala Gly Lys Arg Gly Gly Lys Pro Ser  
 AACGAGGATTGCTGAAGCTCGCGCCGGCGCCGTGCTGGAGTTGGCGCAAGCGTGGCGCAAGCCGTCG  
 1291 1301 1311 1321 1331 1341 1351  
 Pro Lys Gln Arg Asp Ile Ser Leu Arg Leu Val Glu Leu Ala Arg Ala Gly Asn Arg Val Leu Arg Leu Lys Gly  
 CCGAACGAGCGCAGCATCTCGCTCGCTCGAACCTCGCGCCGGCAACCGGGTGCTGGCCCTCAAAGGC  
 1366 1376 1386 1396 1406 1416 1426  
 Gly Asp Pro Phe Val Phe Gly Arg Gly Gly Glu Glu Ala Leu Thr Leu Val Glu His Gln Val Pro Phe Arg Ile  
 GGCGATCCCTTCGTCTCGGTCGCGGGCGAGGAGGCGCTGACGCTGGTCGAACACCAGGTGCCGTTCCGAATC  
 1441 1451 1461 1471 1481 1491 1501  
 Val Pro Gly Ile Thr Ala Gly Ile Gly Gly Leu Ala Tyr Ala Gly Ile Pro Val Thr His Arg Glu Val Asn His  
 GTGCCCGGCATCACCGCCGGTATCGGCGGGCTTGCCTATGCCGGCATCCCGTGAACCCATCGCGAGGTCAACCAC  
 1516 1526 1536 1546 1556 1566 1576  
 Ala Val Thr Phe Leu Thr Gly His Asp Ser Ser Gly Leu Val Pro Asp Arg Ile Asn Trp Gln Gly Ile Ala Ser  
 GCGGTCACCTTCTGACTGGCCATGATTCCCTCGGCTGGTGGCGATCGCATCAACTGGCAGGGCATGCCAGC  
 1591 1601 1611 1621 1631 1641 1651  
 Gly Ser Pro Val Ile Val Met Tyr Met Ala Met Lys His Ile Gly Ala Ile Thr Ala Asn Leu Ile Ala Gly Gly  
 GGCTCGCCTGTCATCGTACATGGCAGTAACATATCGCGCGATCACCGCCAACCTCATGCCGGCGGC  
 1666 1676 1686 1696 1706 1716 1726  
 Arg Ser Pro Asp Glu Pro Val Ala Phe Val Cys Asn Ala Ala Thr Pro Gln Gln Ala Val Leu Glu Thr Thr Leu  
 CGCTCGCCGGACGAACCGGTGCGCTTCGTCTGCACACGCCGCGACGCCGAGCAGGCGGTGCTGGAAACGACGCTT  
 1741 1751 1761 1771 1781 1791 1801  
 Ala Arg Ala Glu Ala Asp Val Ala Ala Ala Gly Leu Glu Pro Pro Ala Ile Val Val Val Gly Glu Val Val Arg  
 GCGCGTGCAGAGGCCGATGTTGCGCGGCAGGGCTGGAGGCCGGCGATCGTCGTCGTCGGCGAGGTGGTGGCG  
 1816 1826 1836 1846 1856 1866 1876  
 Leu Arg Ala Ala Leu Asp Trp Ile Gly Ala Leu Asp Gly Arg Lys Leu Ala Ala Asp Pro Phe Ala Asn Arg Ile  
 CTGCGCGCAGCGCTCGACTGGATCGGCCGCTGGACGGCGCAAGCTTGCCTGCCGACCCGTTGCCAATCGCATT  
 1891 1901 1911 1921 1931 1941 1951  
 Leu Arg Asn Pro Ala \*\*\*  
 CTCAGGAACCCGGCATGA  
 1966 1976 1986 1996 2006 2016 2026

FIGURE 15.1

41/151

NOM = COBA

PREMIER RESIDU = 1  
DERNIER RESIDU = 280

	NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	8	2.86
2	LEU	L	31	11.07
3	ILE	I	16	5.71
4	MET	M	4	1.43
5	VAL	V	27	9.64
6	SER	S	8	2.86
7	PRO	P	19	6.79
8	THR	T	10	3.57
9	ALA	A	41	14.64
10	TYR	Y	2	0.71
11	*	*	0	0.00
12	HIS	H	7	2.50
13	GLN	Q	6	2.14
14	ASN	N	9	3.21
15	LYS	K	8	2.86
16	ASP	D	15	5.36
17	GLU	E	13	4.64
18	CYS	C	2	0.71
19	TRP	W	3	1.07
20	ARG	R	19	6.79
21	GLY	G	32	11.43
22	-	-	0	0.00

RESIDUS = 280  
 POIDS MOLECULAIRE = 29234.  
 INDEX DE POLARITE (%) = 34.  
 POINT ISOELECTRIQUE = 7.51  
 DO 260 (1mg/ml) = 0.464 DO 280 (1mg/ml) = 0.652

## PROFIL D'HYDROPHILICITE DE LA PROTEINE COBA

COBA DE 1 A 280

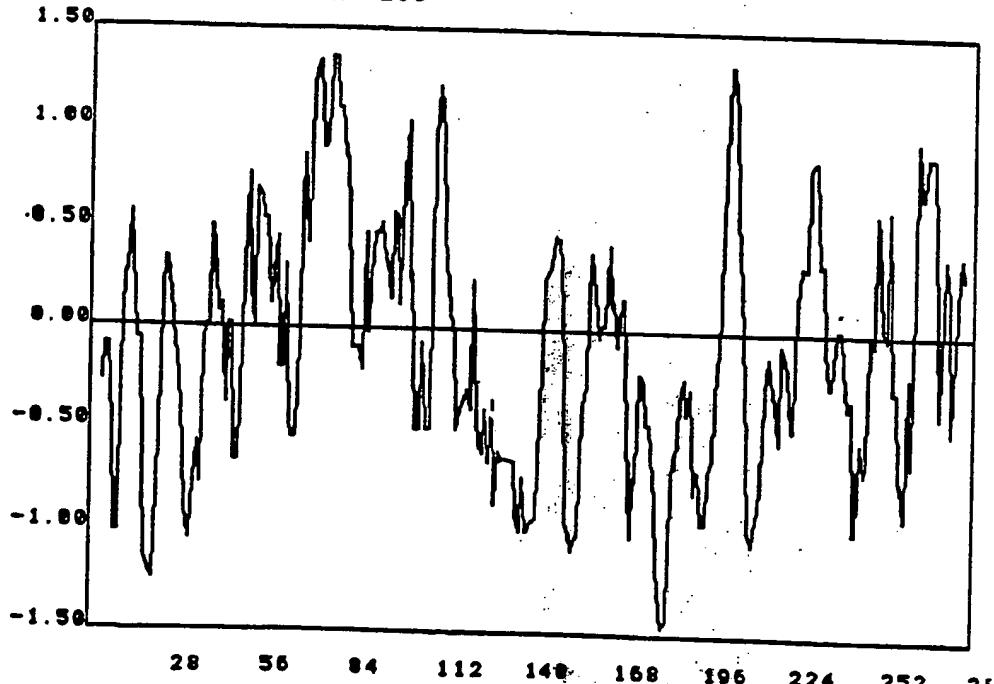


FIGURE 15.2

42 / 351

GENE cobB ET PROTEINE COBBSEQUENCE DU FRAGMENT Clai-HindIII-HindIII DE 5396 BP, DE 1980 A 3281

MetSerGlyLeuLeuIleAlaAlaProAlaSerGlySerGlyLysThrThrValThrLeuGlyLeuMetArgAla  
 ATGAGCGGATTGCTGATTGCCGCACCCGGTCCGGCAAGACGACGGTGACGCTCGGGCTGATGCGCGCC  
 1980 1990 2000 2010 2020 2030 2040  
 LeuLysArgArgGlyValAlaIleAlaProGlyLysAlaGlyProAspTyrIleAspProAlaPheHisAlaAla  
 CTGAAGAGGCGCGCGTGGCGATCGCGCCGGCAAGGGCGGGCCGGACTATATCGATCCGCTTCACGCGGGC  
 2055 2065 2075 2085 2095 2105 2115  
 AlaThrGlyGluProCysPheAsnTyrAspProTrpAlaMetArgProGluLeuLeuAlaAsnAlaSerHis  
 GCGACCGCGAGGCCCTGCTCAACTACGACCCCTGGCGATGCCCGGAACTGCTGCTGCCAATGCGTCGGCAT  
 2130 2140 2150 2160 2170 2180 2190  
 ValAlaSerGlyGlyArgThrLeuIleValGluAlaMetMetGlyLeuHisAspGlyAlaAlaAspGlySerGly  
 GTGGCCTCCGGCGGGCGCACATTGATCGCGAGGCGATGATGGGACTGCGATGACGGTGCCTGCCGACGGCTCGGG  
 2205 2215 2225 2235 2245 2255 2265  
 ThrProAlaAspLeuAlaAlaThrLeuAsnLeuAlaValIleLeuValAlaAspCysAlaArgMetSerGlnSer  
 ACGCCAGCGGACCTCGCCGCGACGCTGAACTTGCCTGGTCAATTCTGGTGGTCATTGCGCCCCGATGTCCCAGTC  
 2280 2290 2300 2310 2320 2330 2340  
 ValAlaAlaLeuValArgGlyTyrAlaAspHisArgAspAspIleArgValValGlyValIleLeuAsnLysVal  
 GTTGCCGCCCTCGCGCGCTATGCGGATCATCGCGACGATATCCGGGTGGCGTCATCCTAACAAAGGTC  
 2355 2365 2375 2385 2395 2405 2415  
 GlySerAspArgHisGluMetMetLeuArgAspAlaLeuGlyLysValArgMetProValPheGlyValLeuArg  
 GGCAGCGATCGGCATGAAATGATGCTGCGCGATGCGCTCGGCAAGGTGCGCATGCGCTGTCTGGCGTGCTCCGG  
 2430 2440 2450 2460 2470 2480 2490  
 GlnAspSerAlaLeuGlnLeuProGluArgHisLeuGlyLeuValGlnAlaGlyGluHisSerAlaLeuGluGly  
 CAGGACAGCGCATTGCAACTGCCGGAGGCCATCTCGGGCTCGCAGGGCGAACACTCAGCGCTTGAGGGC  
 2505 2515 2525 2535 2545 2555 2565  
 PheIleGluAlaAlaAlaAlaArgValGluAlaAlaCysAspLeuAspAlaIleArgLeuIleAlaThrIlePhe  
 TTTCATCGAGGCAGGCCGCGCGGGTCGAGGCTGCGCTGCGATCTCGACGCCATCCGCGATCGCGACGATTTC  
 2580 2590 2600 2610 2620 2630 2640  
 ProGlnValProAlaAlaAlaAspAlaGluArgLeuArgProLeuGlyGlnArgIleAlaValAlaArgAspIle  
 CCGCAGGTGCCGCCGGCGCGATGCCGAGCGTTGCGGCCGCTCGGTGAGCGCATCGCGTCGCGCGATATC  
 2655 2665 2675 2685 2695 2705 2715  
 AlaPheAlaPheCysTyrGluHisLeuLeuTyrGlyTrpArgGlnGlyGlyAlaGluIleSerPhePheSerPro  
 GCCTTGCCCTCTGCTACGAGCACCTGCTTACGGCTGGCGCAAGGGCGCGGAGATTCCCTTCTCGCCG  
 2730 2740 2750 2760 2770 2780 2790  
 LeuAlaAspGluGlyProAspAlaAlaAspAlaValTyrLeuProGlyGlyTyrProGluLeuHisAlaGly  
 CTCGCCACGAGGGGCCGGATGCGGCAGCCGATGCCGCTATCTCCGGGGTTATCCGGAGCTGCATGCGGGG  
 2805 2815 2825 2835 2845 2855 2865  
 GlnLeuSerAlaAlaAlaArgPheArgSerGlyMetHisSerAlaAlaGluArgGlyAlaArgIlePheGlyGlu  
 CAGCTGAGCGCCGCCGCCCCGATTCCGCGATGCCGAGCGGCCGAGCGCGCCGCATCTCGCGAG  
 2880 2890 2900 2910 2920 2930 2940  
 CysGlyGlyTyrMetValLeuGlyGluGlyLeuValAlaAlaAspGlyThrArgTyrAspMetLeuGlyLeuLeu  
 TGCGGCCGGCTATATGGTGCTCGCGAAGGGCTTGCGCTGCCGATGGCACACGCTACGACATGCTCGGCCGCTGCTG  
 2955 2965 2975 2985 2995 3005 3015  
 ProLeuValThrSerPheAlaGluArgArgHisLeuGlyTyrArgArgValValProValAspAsnAlaPhe  
 CCGCTCGTAACCAGTTTGCCGAGCGCAGGCCACCTCGGCTATGCCGCGCTGCGCTGTCGACAACGCC  
 3030 3040 3050 3060 3070 3080 3090  
 PheAspGlyProMetThrAlaHisGluPheHisTyrAlaThrIleValAlaGluGlyAlaAlaAspArgLeuPhe  
 TTCGATGGACCCATGACGGCGCAGGAATTCCACTATGCGACCATCGTCGCCGAAGGGCGGCCGATCGGCTGTTT  
 3105 3115 3125 3135 3145 3155 3165  
 AlaValSerAspAlaAlaGlyGluAspLeuGlyGlnAlaGlyLeuArgArgGlyProValAlaGlySerPheMet  
 GCGGTCAAGCGACGCCGCCGGCGAGGATCTCGGCCAGGCAGGGCCCTCCGGCGCGGCCCTGCGCCGGTTCC  
 3180 3190 3200 3210 3220 3230 3240  
 HisLeuIleAspValAlaGlyAlaAla\*\*\*  
 CATCTGATCGACGTCGCAAGGTGCTGCATGA  
 3255 3265 3275 3285 3295 3305 3315

43/151

NOM = COBB

PREMIER RESIDU = 1  
DERNIER RESIDU = 434

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	17	3.92	2500.19	5.47
2	LEU	L	45	10.37	5088.60	11.14
3	ILE	I	17	3.92	1922.36	4.21
4	MET	M	14	3.23	1834.56	4.02
5	VAL	V	31	7.14	3071.17	6.72
6	SER	S	19	4.38	1653.57	3.62
7	PRO	P	21	4.84	2038.05	4.46
8	THR	T	12	2.76	1212.60	2.65
9	ALA	A	76	17.51	5399.04	11.82
10	TYR	Y	11	2.53	1793.66	3.93
11	*	*	0	0.00	0.00	0.00
12	HIS	H	14	3.23	1918.84	4.20
13	GLN	Q	9	2.07	1152.54	2.52
14	ASN	N	5	1.15	570.20	1.25
15	LYS	K	5	1.15	640.45	1.40
16	ASP	D	28	6.45	3220.84	7.05
17	GLU	E	21	4.84	2709.84	5.93
18	CYS	C	5	1.15	515.05	1.13
19	TRP	W	2	0.46	372.16	0.81
20	ARG	R	34	7.83	5307.40	11.62
21	GLY	G	48	11.06	2736.96	5.99
22	-	-	0	0.00	0.00	0.00

RESIDUS = 434  
 POIDS MOLECULAIRE = 45676.  
 INDEX DE POLARITE (%) = 34.  
 POINT ISOELECTRIQUE = 6.47  
 DO 260 (1mg/ml) = 0.351 DO 280 (1mg/ml) = 0.529

PROFIL D'HYDROPHILICITE DE LA PROTEINE  
COBB DE 1 A 434

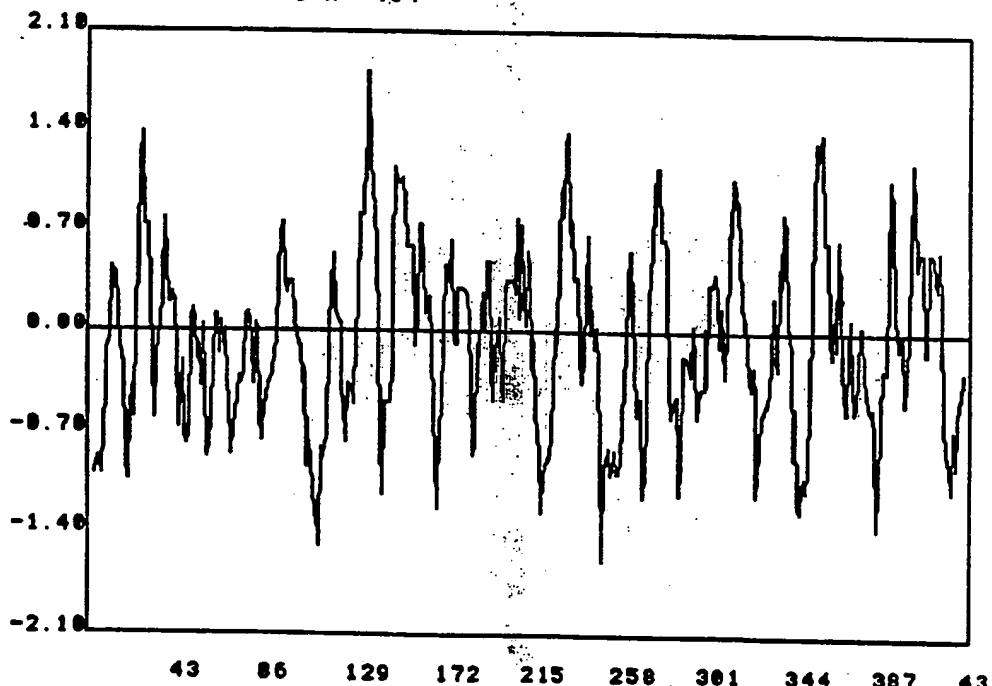


FIGURE 15.4

44/151

GENE cobC ET PROTEINE COBCSEQUENCE DU FRAGMENT Clai-HindIII-HindIII-HindIII DE 5396 BP, DE 13281 A 4279

MetSerAlaProIleValHisGlyGlyIleThrGluAlaAlaAlaArgTyrGlyGlyArgProGluAspTrp  
 ATGAGCGCACCGATCGTCATGGTGGCGGCATCACCGAGGCCGAGCGCGCTATGGCGGCCGCGCTGAAGACTGG  
 3281 3291 3301 3311 3321 3331 3341  
 LeuAspLeuSerThrGlyIleAsnProCysProValAlaLeuProAlaValProGluArgAlaTrpHisArgLeu  
 CTCGATCTGTCGACCGGCATCAATCCATGCCCGTCCCTGCCCGGGTCCCTGAGCGCGCCCTGGCACCGGCTG  
 3356 3366 3376 3386 3396 3406 3416  
 ProAspArgGlnThrValAspAspAlaArgSerAlaAlaAlaAspTyrTyrArgThrAsnGlyValLeuProLeu  
 CGGGATCGGCAGACGGTAGATGATGCGCGAGCGCCGCGACTACTACCGCACCAACGGCGTGCTGCCTTG  
 3431 3441 3451 3461 3471 3481 3491  
 ProValProGlyThrGlnSerValIleGlnLeuLeuProArgLeuAlaProAlaAsnArgHisValAlaIlePhe  
 CCGGTGCCGGGCACCCAGTCGGTGTCCAGCTCCTGCCACGTCTGCTCCGGCCAACAGGCACGTCGGATTTC  
 3506 3516 3526 3536 3546 3556 3566  
 GlyProThrTyrGlyGluTyrAlaArgValLeuGluAlaAlaGlyPheAlaValAspArgValAlaAspAlaAsp  
 GGGCCGACCTATGGCGACTATGCCCGCGTCTGAAAGGGCCGGCTTGCTGTCGATCGCGCGGATGCCGAC  
 3581 3591 3601 3611 3621 3631 3641  
 AlaLeuThrAlaGluHisGlyLeuValIleValValAsnProAsnAsnProThrGlyArgAlaLeuAlaProAla  
 GCGCTCACGGCCGAAACATGGGCTTGTCACTCGCTCAACCCAAACCCGACC GGCGCCTGGCGCCGGCG  
 3656 3666 3676 3686 3696 3706 3716  
 GluLeuLeuAlaIleAlaAlaArgGlnLysAlaSerGlyGlyLeuLeuLeuValAspGluAlaPheGlyAspLeu  
 GAGCTTCTGGCGATGCCGCAAGGCAGAAGGCGAGCGCGGACTGCTGCTGGTCGATGAGGCCTTCGGCGATCTT  
 3731 3741 3751 3761 3771 3781 3791  
 GluProGlnLeuSerValAlaGlyHisAlaSerGlyGlnGlyAsnLeuIleValPheArgSerPheGlyLysPhe  
 GAGCCGCAACTGAGTGTGCGTGGTCACGCGTCAGGGCAAGGCAACCTCATCGCTTCCGCTCTCGGCAAGTTC  
 3806 3816 3826 3836 3846 3856 3866  
 PheGlyLeuAlaGlyLeuArgLeuGlyPheValValAlaThrGluProValLeuAlaSerPheAlaAspTrpLeu  
 TTGGCCTTGGGGCTGCGCCTCGGCTTCGTCGACCGAGCCAGTGGCTGATCCTTGCCGATTGGCTC  
 3881 3891 3901 3911 3921 3931 3941  
 GlyProTrpAlaValSerGlyProAlaLeuThrIleSerLysAlaLeuMetGlnGlyAspThrLysAlaIleAla  
 GGTCCCTGGGCTGTCCTCCGGCCCGCGTIGACGATCTCGAAAGCGCTGATGCAAGGGCGATACGAAGGCGATCGCG  
 3956 3966 3976 3986 3996 4006 4016  
 AlaGlyIleLeuGluArgArgAlaGlyLeuAspAlaAlaLeuAspGlyAlaGlyLeuAsnArgIleGlyGlyThr  
 GCGGGCATCCTCGAGCGTCGCCGCCCTCGATGCCCTCTCGATGGGGCAGGGCTAACCGTATGGCGGCACG  
 4031 4041 4051 4061 4071 4081 4091  
 GlyLeuPheValLeuValGluHisProArgAlaAlaLeuGlnGluArgLeuCysGluAlaHisIleLeuThr  
 GGGCTATTGCGTGGTCGAGGACATCCCAGGGCAGCTCTGCTGCAGGAGCGGCTCTGCGAGGGCCATATTCTCAGG  
 4106 4116 4126 4136 4146 4156 4166  
 ArgLysPheAspTyrAlaProThrTrpLeuArgValGlyLeuAlaProAspAlaAlaGlyAspArgArgLeuAla  
 CGCAAGTTGACTATGCCCGACCTGGCTCAGGGTGGCTTGCGCCCTGACGCCGCTGGTGACCGACGGCTGGCG  
 4181 4191 4201 4211 4221 4231 4241  
 AspAlaLeuAlaArgMetGluLeu\*\*\*  
 GACGCGCTTGGCCGATGGAGCTCTGA  
 4256 4266 4276 4286 4296 4306 4316

FIGURE 15.5

45/151

NOM = COBC

PREMIER RESIDU = 1  
DERNIER RESIDU = 333

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	11	3.30	1617.77	4.62
2	LEU	L	43	12.91	4862.44	13.90
3	ILE	I	13	3.90	1470.04	4.20
4	MET	M	3	0.90	393.12	1.12
5	VAL	V	24	7.21	2377.68	6.79
6	SER	S	11	3.30	957.33	2.74
7	PRO	P	23	6.91	2232.15	6.38
8	THR	T	14	4.20	1414.70	4.04
9	ALA	A	56	16.82	3978.24	11.37
10	TYR	Y	6	1.80	978.36	2.80
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	2.10	959.42	2.74
13	GLN	Q	8	2.40	1024.48	2.93
14	ASN	N	8	2.40	912.32	2.61
15	LYS	K	5	1.50	640.45	1.83
16	ASP	D	19	5.71	2185.57	6.25
17	GLU	E	15	4.50	1935.60	5.53
18	CYS	C	2	0.60	206.02	0.59
19	TRP	W	5	1.50	930.40	2.66
20	ARG	R	25	7.51	3902.50	11.15
21	GLY	G	35	10.51	1995.70	5.70
22	-	-	0	0.00	0.00	0.00

RESIDUS = 333

POIDS MOLECULAIRE = 34992.

INDEX DE POLARITE (%) = 34.

POINT ISOELECTRIQUE = 6.72

DO 260 (1mg/ml) = 0.670 DO 280 (1mg/ml) = 0.998

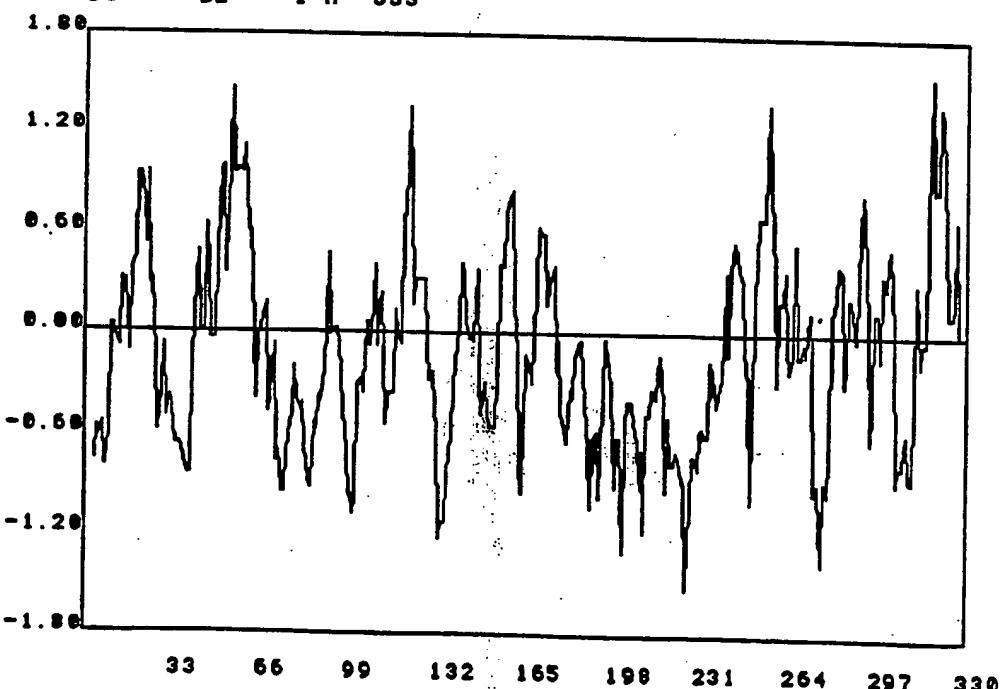
PROFIL D'HYDROPHILICITE DE LA PROTEINE  
COBC DE 1 A 333

FIGURE 15.6

46 / 151

GENE cobD ET PROTEINE COBDSEQUENCE DU FRAGMENT ClaI-HindIII-HindIII-HindIII DE 5396 BP, DE 4284 A 5252

MetSerGluThrIleLeuLeuIleLeuAlaLeuAlaLeuValIleAspArgValValGlyAspProAspTrpLeu  
 GTGTGGAGACGATCCTGCTCATTCGCGCTGGCGTGGTGAATCGACCCGCGTGTGCGCGATCCGGACTGGCTC  
 4284 4294 4304 4314 4324 4334 4344  
 TrpAlaArgValProHisProValValPhePheGlyLysAlaIleGlyPhePheAspAlaArgLeuAsnArgGlu  
 TGGGCGCGCGTGCCTGCATCCGGTCGTGTTTCGGCAAGGCCATCGGCTTTTCGACGCCGCGCTGAACCGGGAG  
 4359 4369 4379 4389 4399 4409 4419  
 AspLeuGluAspSerAlaArgLysPheArgGlyValValAlaIleLeuLeuLeuGlyIleSerAlaTrpPhe  
 GACCTCGAGGATAGCGCGCACAATTCTGTCGCGATCCTTGTTGCTGGCATCAGCGCCTGGTTC  
 4434 4444 4454 4464 4474 4484 4494  
 GlyHisLeuLeuHisArgLeuPheAlaValLeuGlyProLeuGlyPheLeuLeuGluAlaValLeuValAlaVal  
 GGCCATCTGCTGCATCGCTGTCGCGCTCGGACCGCTCGGCTTCTGCTCGAGGCCGTTCTGGTCGCGGTC  
 4509 4519 4529 4539 4549 4559 4569  
 PheLeuAlaGlnLysSerLeuAlaAspHisValArgArgValAlaGlyGlyLeuArgGlnGlyGlyLeuGluGly  
 TTCCTGGCACAGAAGAGCCTCGCCGATCACGTGCGTGGCGGGGCTTGCACAGGGCGGGCTGGAAAGGC  
 4584 4594 4604 4614 4624 4634 4644  
 GlyArgAlaAlaValSerMetIleValGlyArgAspProLysThrLeuAspGluProAlaValCysArgAlaAla  
 GGGCGTGCCTGCCGCGTGTGATCGTGGTGCCTACCAAAGACGCTCGACGAGCCGGCTGCCGTGCCGCG  
 4659 4669 4679 4689 4699 4709 4719  
 IleGluSerLeuAlaGluAsnPheSerAspGlyValValAlaProAlaPheTrpTyrAlaValAlaGlyLeuPro  
 ATCGAAAGCCTGCCGAGAAATTCTCCGACGGCGTGTGGCGCCCTCTGGTACGCCGTTGCCGGCTGCCG  
 4734 4744 4754 4764 4774 4784 4794  
 GlyLeuLeuAlaTyrLysMetLeuAsnThrAlaAspSerMetIleGlyHisLysSerProLysTyrLeuHisPhe  
 GGGCTTCTTGCCTACAAGATGCTAACACCGCCGATTGATGCCAACAGTCGCCAACATCTGCACTTC  
 4809 4819 4829 4839 4849 4859 4869  
 GlyTrpAlaSerAlaArgLeuAspAspLeuAlaAsnLeuProAlaAlaArgLeuSerIleLeuLeuIleSerAla  
 GGCTGGGCCTCGGCCGACTCGACGATCTGCCAACCTGCCGGCAGCGAGGGCTCTGATCCTTGTATCTCAGCC  
 4884 4894 4904 4914 4924 4934 4944  
 GlyAlaLeuIleHisArgGlyAlaSerAlaAlaIlysAspAlaLeuThrValAlaLeuArgAspHisGlyLeuHis  
 GGTGCGCTGATCCATCGTGGCGCCAGCGCCGCCAGGATGCGCTGACCGTGGCCCTCGCGACCATGCCCTGCAC  
 4959 4969 4979 4989 4999 5009 5019  
 ArgSerProAsnSerGlyTrpProGluAlaAlaMetAlaGlyAlaLeuAspLeuGlnLeuAlaGlyProArgIle  
 CGCTGCCGAACTCCGGCTGCCGGAAAGCGGCCATGCCGGCGCTCGATCTGCAGCTTGCCGGTCCGGATC  
 5034 5044 5054 5064 5074 5084 5094  
 TyrGlyGlyValLysValSerGluProMetIleAsnGlyProGlyArgAlaValAlaThrSerGluAspIleAsp  
 TATGGCGCGTCAAGGTCAAGCGAACCTATGATCAACGGTCCGGCGAGCGGTTGCAACAAGCGAACATCGAC  
 5109 5119 5129 5139 5149 5159 5169  
 AlaGlyIleAlaValPheTyrGlyAlaCysThrValMetAlaGlyPheValLeuAlaIleAlaMetIle\*\*\*  
 GCCGGTATTGCTGTATTTATGGCGCTGTACGGTATGGCCGGTTGTTCTGCAATCGCAATGATTGA  
 5184 5194 5204 5214 5224 5234 5244

FIGURE 15.7

47/151

NOM = COBD

PREMIER RESIDU = 1  
DERNIER RESIDU = 323

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	14	4.33	2058.98	6.02
2	LEU	L	45	13.93	5088.60	14.89
3	ILE	I	18	5.57	2035.44	5.96
4	MET	M	8	2.48	1048.32	3.07
5	VAL	V	27	8.36	2674.89	7.83
6	SER	S	17	5.26	1479.51	4.33
7	PRO	P	15	4.64	1455.75	4.26
8	THR	T	6	1.86	606.30	1.77
9	ALA	A	51	15.79	3623.04	10.60
10	TYR	Y	5	1.55	815.30	2.39
11	*	*	0	0.00	0.00	0.00
12	HIS	H	9	2.79	1233.54	3.61
13	GLN	Q	3	0.93	384.18	1.12
14	ASN	N	6	1.86	684.24	2.00
15	LYS	K	9	2.79	1152.81	3.37
16	ASP	D	18	5.57	2070.54	6.06
17	GLU	E	11	3.41	1419.44	4.15
18	CYS	C	2	0.62	206.02	0.60
19	TRP	W	6	1.86	1116.48	3.27
20	ARG	R	20	6.19	3122.00	9.14
21	GLY	G	33	10.22	1881.66	5.51
22	-	-	0	0.00	0.00	0.00

RESIDUS = 323

POIDS MOLECULAIRE = 34175.

INDEX DE POLARITE (%) = 31.

POINT ISOELECTRIQUE = 8.00

DO 260 (1mg/ml) = 0.789 DO 280 (1mg/ml) = 1.150

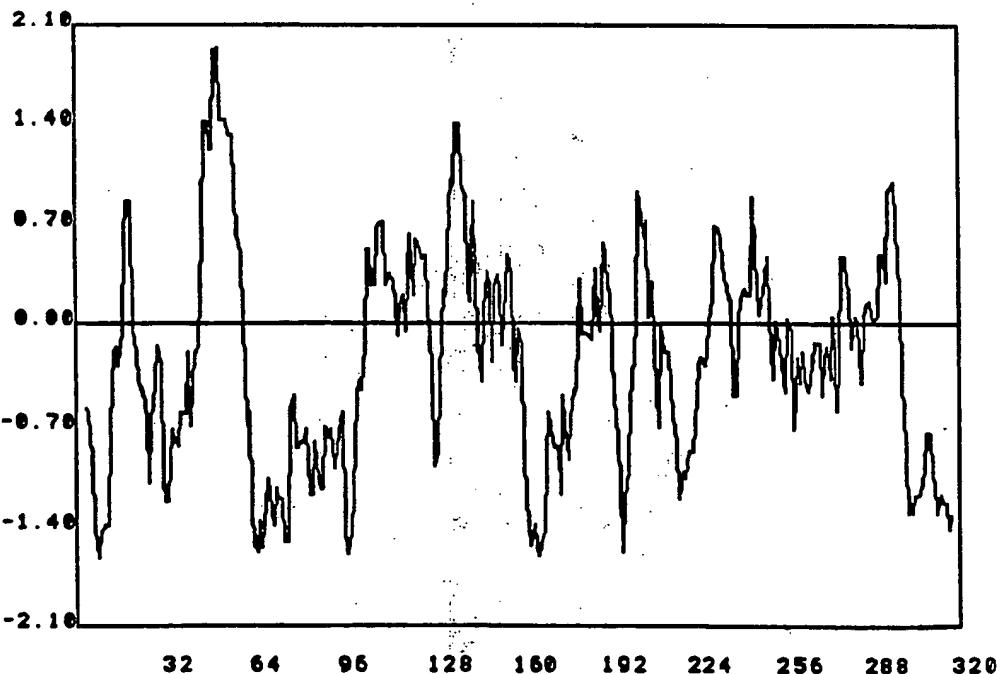
PROFIL D'HYDROPHILICITE DE LA PROTEINE  
COBD DE 1 A 323

FIGURE 15.8

48 / 151

## GENE COBE ET PROTEINE COBE

SEQUENCE DU FRAGMENT Clal-HindIII-HindIII DE 5396 BP, DE 549 A 1010

Met Pro Ser Gly Gln His Ser Ala Glu Gln Thr Thr Lys <sup>549</sup>  
 ATGCCATCGGGCCAACACTCTGCACAGACGAAAGCAGGAGCCGGCTGGTGCCTGGCTCGGCTGCGAGCGT  
 559 569 579 589 599 609  
 Arg Thr Pro Ala Glu Glu Val Ile Ala Leu Ala Glu Arg Ala Leu Ala Asp Ala Gly Val Ala Pro Gly Asp Leu  
 CGCACGCCGCCGAAGAGGTGATGCCCTTGCCGAGCGTGCCTGCCGATGCCGGTGTGCGCCGGCGATCTG  
 624 634 644 654 664 674 684  
 Arg Leu Val Ala Ser Leu Asp Ala Arg Ala Glu Glu Pro Ala Ile Leu Ala Ala Ala Glu His Phe Ala Val Pro  
 CGGCTGGTCGCCTCGCTCGATGCTCGCCGAGGAGCCGGCGATCCCTGGCGCCGCTCAGCATTTCGCGGGTCCG  
 699 709 719 729 739 749 759  
 Ala Ala Phe Tyr Asp Ala Ala Thr Leu Glu Ala Glu Ala Ser Arg Leu Ala Asn Pro Ser Glu Ile Val Phe Ala  
 GCCGCCTTCTACGATGCCAACGCTCGAAGCCGAAGCTCCCGCTGCCAACCCGTCCGAGATCGTCTTGCC  
 774 784 794 804 814 824 834  
 Tyr Thr Gly Cys His Gly Val Ala Glu Gly Ala Ala Leu Val Gly Ala Gly Arg Glu Ala Val Leu Ile Val Gln  
 TACACGGGTTGTCATGGCGTTGCCGAGGGTGCAGCGCTCGTCGGCGCCGGTCGCGAAGCCGTGCTGATTGTGCAG  
 849 859 869 879 889 899 909  
 Lys Ile Val Ser Ala His Ala Thr Ala Ala Leu Ala Glu Pro Ala Thr Leu Arg Ala Glu Lys Arg Ile Gln Ala  
 AAGATCGTCTCCGCCATGCGACGGCCGACTTGCCGGCCGGCGACCTTGCGCGCCGAAAAGCGCATCCAGGCC  
 924 934 944 954 964 974 984  
 Ala Glu Ala Val \*\*\*  
 GCGGAGGGCTGTCTGA  
 999 1009 1019 1029 1039 1049 1059

FIGURE 15.9

49/151

NOM = COBE

PREMIER RESIDU = 1  
DERNIER RESIDU = 154

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	3	1.95	441.21	2.85
2	LEU	L	15	9.74	1696.20	10.96
3	ILE	I	6	3.90	678.48	4.38
4	MET	M	1	0.65	131.04	0.85
5	VAL	V	12	7.79	1188.84	7.68
6	SER	S	6	3.90	522.18	3.37
7	PRO	P	7	4.55	679.35	4.39
8	THR	T	7	4.55	707.35	4.57
9	ALA	A	41	26.62	2912.64	18.82
10	TYR	Y	2	1.30	326.12	2.11
11	*	*	0	0.00	0.00	0.00
12	HIS	H	4	2.60	548.24	3.54
13	GLN	Q	5	3.25	640.30	4.14
14	ASN	N	1	0.65	114.04	0.74
15	LYS	K	3	1.95	384.27	2.48
16	ASP	D	4	2.60	460.12	2.97
17	GLU	E	13	8.44	1677.52	10.84
18	CYS	C	2	1.30	206.02	1.33
19	TRP	W	0	0.00	0.00	0.00
20	ARG	R	9	5.84	1404.90	9.08
21	GLY	G	13	8.44	741.26	4.79
22	-	-	0	0.00	0.00	0.00

RESIDUS = 154

POIDS MOLECULAIRE = 15478.

INDEX DE POLARITE (%) = 34.

POINT ISOELECTRIQUE = 5.61

DO 260 (1mg/ml) = 0.113 DO 280 (1mg/ml) = 0.154

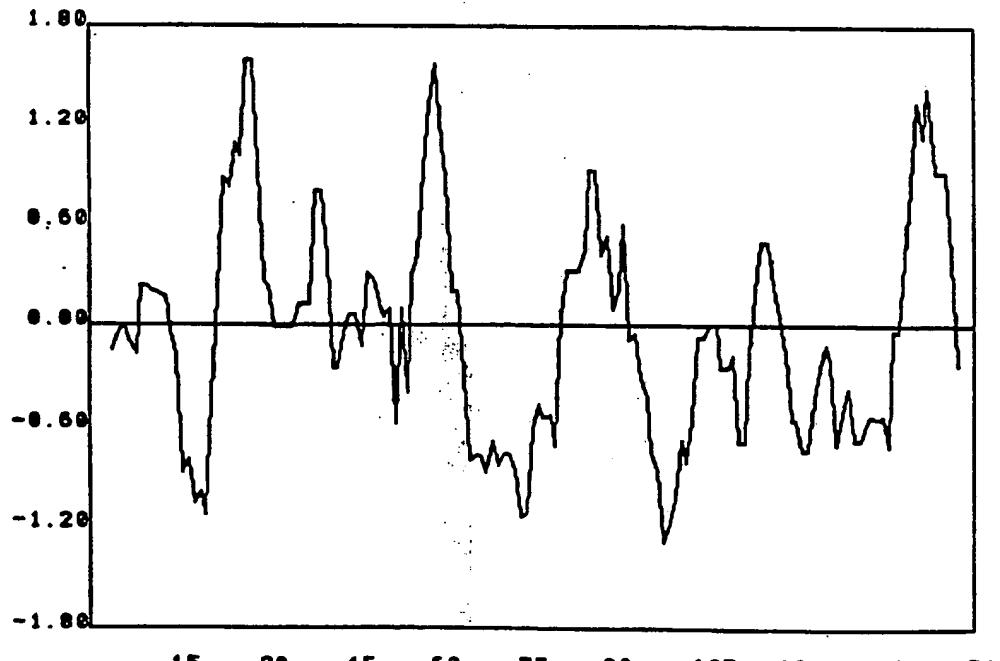
PROFIL D'HYDROPHILICITE DE LA PROTEINE  
COBE DE 1 A 154

FIGURE 15.10.

50 / 151

GENE cobF ET PROTEINE COBF

SEQUENCE DU FRAGMENT DE 8753 BP, DE 736 A 1521

Met Ala Glu Ala Gly Met Arg Lys Ile Leu Ile Ile Gly Ser Gly Asn Pro Glu His Met Thr Val Gln  
 ATGGCGGAGGCAGGGCATGCGCAAAATTCTGATCATCGGCATCGGTTGGCAATCCGAACACATGACCGTGCAG  
 736      746      756      766      776      786      796  
 Ala Ile Asn Ala Leu Asn Cys Ala Asp Val Leu Phe Ile Pro Thr Lys Gly Ala Lys Lys Thr Glu Leu Ala Glu  
 GCGATCAACGCGCTGAAGTGGCCGACGTGCTCTTATCCCACCAAGGGAGCGAAGAAGACCGAGCTTGCCGAA  
 811      821      831      841      851      861      871  
 Val Arg Arg Asp Ile Cys Ala Arg Tyr Val Thr Arg Lys Asp Ser Arg Thr Val Glu Phe Ala Val Pro Val Arg  
 GTGCGCCGCGACATCTGCGCCGCTACGTCACGCGCAAGGACAGCCGCACCGTCGAGTTCGCGGTGCCCGTGC  
 886      896      906      916      926      936      946  
 Arg Thr Glu Gly Val Ser Tyr Asp Gly Ser Val Asp Asp Trp His Ala Gln Ile Ala Gly Ile Tyr Glu Ala Leu  
 CGCACCGAAGGCGTCAGCTATGACGGCAGCGTCGATGACTGGCACGCCAGATCGCTGGGATTACGAAGCGCTT  
 961      971      981      991      1001      1011      1021  
 Leu Ser Lys Glu Leu Gly Glu Glu Gly Thr Gly Ala Phe Leu Val Trp Gly Asp Pro Met Leu Tyr Asp Ser Thr  
 CTATCGAAGGAGTTGGCGAACGGAACTGGCGCTTCTCGCTGGCGACCCGATGCTCTATGACAGCACC  
 1036      1046      1056      1066      1076      1086      1096  
 Ile Arg Ile Val Glu Arg Val Lys Ala Arg Gly Glu Val Ala Phe Ala Tyr Asp Val Ile Pro Gly Ile Thr Ser  
 ATT CGC AT CGT CGAGCGGGTCAAGGCACCGCGTGAGGTGCGCTACGACGT CATTCCGGATCACCAGT  
 1111      1121      1131      1141      1151      1161      1171  
 Leu Gln Ala Leu Cys Ala Ser His Arg Ile Pro Leu Asn Leu Val Gly Lys Pro Val Glu Ile Thr Thr Gly Arg  
 CTG CAGGCGCTTGCGCCAGCCACCGCATTCCGCTGAACCTCGTCGGCAAGCCGGTGGAGATCACACGGGGCGT  
 1186      1196      1206      1216      1226      1236      1246  
 Arg Leu His Glu Ser Phe Pro Glu Lys Ser Gln Thr Ser Val Val Met Leu Asp Gly Glu Gln Ala Phe Gln Arg  
 CGGCTGCACGAAAGCTTCCCGAGAAGGCCAGACCTCGGTGCTCATGCTCGATGGCGAACAGGCCGTTAGCGG  
 1261      1271      1281      1291      1301      1311      1321  
 Val Glu Asp Pro Glu Ala Glu Ile Tyr Trp Gly Ala Tyr Leu Gly Thr Arg Asp Glu Ile Val Ile Ser Gly Arg  
 GTCGAGGACCCGGAGGCGGAGATCTATTGGGGCGCTATCTGGCACGCCGATGAGATCGTCATTCCGGCCG  
 1336      1346      1356      1366      1376      1386      1396  
 Val Ala Glu Val Lys Asp Arg Ile Leu Glu Thr Arg Ala Ala Ala Arg Ala Lys Met Gly Trp Ile Met Asp Ile  
 GTGGCTGAGGTGAAGGACCGGATCCTGAAACGCGGGCGGGCGCCGAAGATGGATGGATCATGGACATC  
 1411      1421      1431      1441      1451      1461      1471  
 Tyr Leu Leu Arg Lys Gly Ala Asp Phe Asp Glu \*\*\*  
 TATCTCCTGCGCAAGGGCGCCGACTTCGACGAGTGAA  
 1486      1496      1506      1516

FIGURE 16.1

51/151

PROTEINE COBF

PREMIER RESIDU = 1  
DERNIER RESIDU = 261

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	7	2.68	1029.49	3.56
2	LEU	L	19	7.28	2148.52	7.43
3	ILE	I	21	8.05	2374.68	8.21
4	MET	M	7	2.68	917.28	3.17
5	VAL	V	22	8.43	2179.54	7.53
6	SER	S	12	4.60	1044.36	3.61
7	PRO	P	9	3.45	873.45	3.02
8	THR	T	14	5.36	1414.70	4.89
9	ALA	A	27	10.34	1918.08	6.63
10	TYR	Y	8	3.07	1304.48	4.51
11	*	*	0	0.00	0.00	0.00
12	HIS	H	4	1.53	548.24	1.90
13	GLN	Q	6	2.30	768.36	2.66
14	ASN	N	4	1.53	456.16	1.58
15	LYS	K	12	4.60	1537.08	5.31
16	ASP	D	16	6.13	1840.48	6.36
17	GLU	E	23	8.81	2967.92	10.26
18	CYS	C	3	1.15	309.03	1.07
19	TRP	W	4	1.53	744.32	2.57
20	ARG	R	21	8.05	3278.10	11.33
21	GLY	G	22	8.43	1254.44	4.34

RESIDUS = 261  
 POIDS MOLECULAIRE = 28927.  
 INDEX DE POLARITE (%) = 43.  
 POINT ISOELECTRIQUE = 5.70  
 DO 260 (1mg/ml) = 0.705 DO 280 (1mg/ml) = 1.097

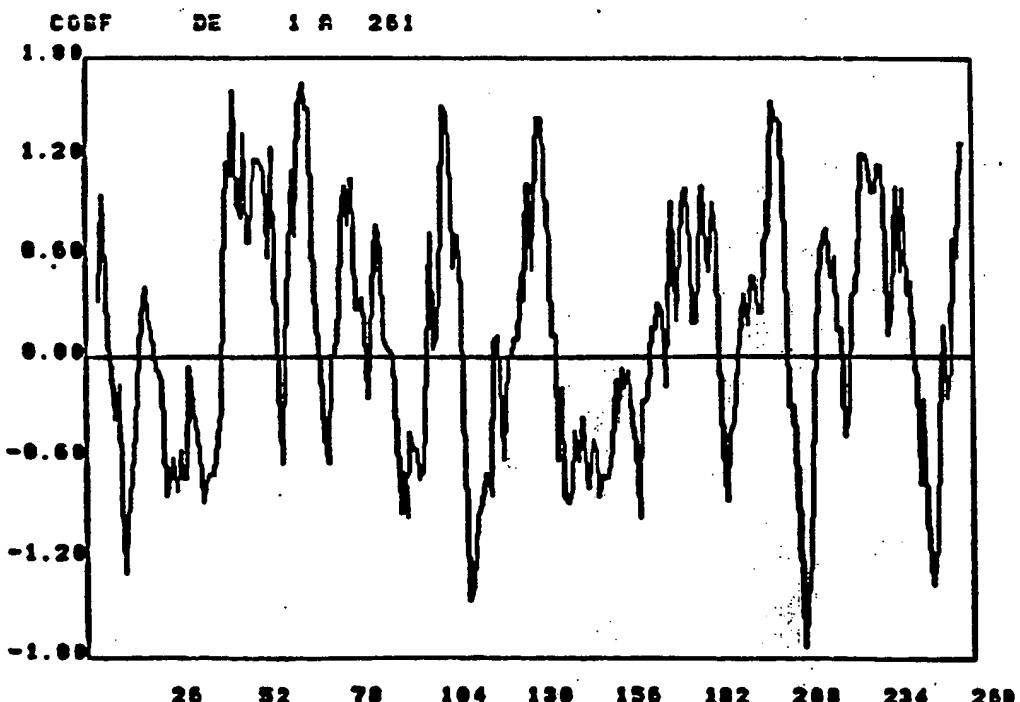


FIGURE 16.2

52/151

GENE cobG ET PROTEINE COBG  
SEQUENCE DU FRAGMENT DE 8753 BP, DE 1620 A 2999

MetThrAspLeuMetThrSerCysAlaLeuProLeuThrGlyAspAlaGlyThrValAlaSerMetArgArgGly  
 ATGACGGATTTGATGACCAGCTGCGCCCTTCCATTGACCGGAGATGCCGGCACCGTCGCTTCGATGCCGCCGGC  
 1620 1630 1640 1650 1660 1670 1680  
 AlaCysProSerLeuAlaGluProMetGlnThrGlyAspGlyLeuLeuValArgValArgProThrAspAspSer  
 GCCTGCCCGTCCTTGGCAGAGCCGATGCAGACCGGGCACGGCCTGCTCGTGGGTGAGGGCAACGGATGACAGC  
 1695 1705 1715 1725 1735 1745 1755  
 LeuThrLeuProLysValIleAlaLeuAlaThrAlaAlaGluArgPheGlyAsnGlyIleIleGluIleThrAla  
 CTGACGCTGCCGAAGGTCAATTGCCCTTGCCACGGCTGCCAGCGCTTCGGCAATGGCATCATCGAGATTACCGCG  
 1770 1780 1790 1800 1810 1820 1830  
 ArgGlyAsnLeuGlnLeuArgGlyLeuSerAlaAlaSerValProArgLeuAlaGlnAlaIleGlyAspAlaGlu  
 CGCGGAAACCTGCAGCTTCGCGGCTGAGCGGGCTCGGTGCCAAGGCTGGCGAGCGATCGCGATGCCGAG  
 1845 1855 1865 1875 1885 1895 1905  
 IleAlaIleAlaGluGlyLeuAlaIleGluValProProLeuAlaGlyIleAspProAspGluIleAlaAspPro  
 ATCGCCATTGCCGAGGGCTCGCGATCGAGGTGCCCTGGCCGGCATCGACCGGACGAGATCGCGATCCG  
 1920 1930 1940 1950 1960 1970 1980  
 ArgProIleAlaThrGluLeuArgGluAlaLeuAspValArgGlnValProLeuLysLeuAlaProLysLeuSer  
 CGGCCGATTGCCACTGAGCTCGTGAAGCGTTGGATGTGCGCCAGGTGCCCTGAAGCTTCACCCAAATTATCC  
 1995 2005 2015 2025 2035 2045 2055  
 ValValIleAspSerGlyGlyArgPheGlyLeuGlyAlaValAlaAspIleArgLeuGlnAlaValSerThr  
 GTCGTCATCGATAAGCGGTGGCGGTTGGTCTCGCGCTGTCGCGACATTGCCCTTCAGGCCGTTTCGACT  
 2070 2080 2090 2100 2110 2120 2130  
 ValAlaGlyValAlaTrpValLeuSerLeuGlyGlyThrSerThrLysAlaSerSerValGlyThrLeuAlaGly  
 GTCGCGGGGGTGGCCTGGGTGCTGCGCTTGGCGCACGTCAACGAAGGCATCGAGCGTCGGGACGTTGGCCGGC  
 2145 2155 2165 2175 2185 2195 2205  
 AsnAlaValValProAlaLeuIleThrIleLeuGluLysLeuAlaSerLeuGlyThrThrMetArgGlyArgAsp  
 AACCGGGTCGTGCCGGCCCTGATCACCATCTCGAGAAACTGGCGAGCCTGGGCACGACGATGCGCGGGCGAT  
 2220 2230 2240 2250 2260 2270 2280  
 LeuAspProSerGluIleArgAlaLeuCysArgCysGluThrSerSerGluArgProAlaAlaProArgSerAla  
 CTGGACCCGTCGAAATCCGCGCCTGTCGCTGTGAGACATCGTCCGAACGCCGGCCTCCGCGTTGGCG  
 2295 2305 2315 2325 2335 2345 2355  
 AlaIleProGlyIleHisAlaLeuGlyAsnAlaAspThrValLeuGlyLeuGlyLeuAlaPheAlaGlnValGlu  
 GCAATACCGGCATTATCGCCTGGGTAAACGCCGACACCGTTCTGGCCTCGGTCTGGCCTTGCTCAGGTGGG  
 2370 2380 2390 2400 2410 2420 2430  
 AlaAlaAlaLeuAlaSerTyrLeuHisGlnValGlnAlaLeuGlyAlaAsnAlaIleArgLeuAlaProGlyHis  
 GCCGCCGCGCTGGCATCCTAACCTGCATCAGGTCCAGGCGCTTGGCGCCAATGCGATCCGGCTTGGCCCCGGGAC  
 2445 2455 2465 2475 2485 2495 2505  
 AlaPhePheValLeuGlyLeuCysProGluThrAlaAlaValAlaGlnSerLeuAlaAlaSerHisGlyPheArg  
 GCCTCTTCGTCCTCGGCCATTGCCCCGAGACCGCGCTGTGGCGCAGAGCCTGGCAGCGTCACACGGTTTCGC  
 2520 2530 2540 2550 2560 2570 2580  
 IleAlaGluGlnAspProArgAsnAlaIleAlaThrCysAlaGlySerLysGlyCysAlaSerAlaTrpMetGlu  
 ATTGCCGAGCAGGATCCGCGCAATGCGATGCCACCTGCGCCGGCAGCAAGGGTTGCCCTCGCGTGGATGGAA  
 2595 2605 2615 2625 2635 2645 2655  
 ThrLysGlyMetAlaGluArgLeuValGluThrAlaProGluLeuLeuAspGlySerLeuThrValHisLeuSer  
 ACCAAGGGCATGGCGAGGCCCTCGTCGAGACGGGCCGGATTGCTCGACGGGTCGCTCACCGTGCATCTCTCC  
 2670 2680 2690 2700 2710 2720 2730  
 GlyCysAlaLysGlyCysAlaArgProLysProSerGluLeuThrLeuValGlyAlaProSerGlyTyrGlyLeu  
 GGCTGCGCCAAGGGCTGCGCCGGCGAAGCCGCTGAACTGACGCTTGTGGTGCAGCATCAGGATACGGGCTT  
 2745 2755 2765 2775 2785 2795 2805  
 ValValAsnGlyAlaAlaAsnGlyLeuProSerAlaIleThrAspGluAsnGlyMetGlySerAlaLeuAlaArg  
 GTCGTAAATGGGCTGCCAATGGCTTGCCAAAGCGCTACACCGATGAGAATGGAATGGGATCCGCCCTGGCCGG  
 2820 2830 2840 2850 2860 2870 2880  
 LeuGlyArgLeuValArgGlnAsnLysAspAlaGlyGluSerAlaGlnSerCysLeuThrArgLeuGlyAlaAla  
 CTCGGCCGGCTGGTGCAGGCAAACAAAGACGCTGGCGAATCGGCGAGTCCTGTCTTACACGGCTCGGAGCTGG  
 2895 2905 2915 2925 2935 2945 2955  
 ArgValSerAlaAlaPheGluGlnGly\*\*\*  
 CGCGTCTCGGCAGCGTTCGAACAGGGATAG  
 2970 2980 2990 3000

FIGURE 16.3

53/151

PROTEINE COBG

PREMIER RESIDU = 1  
DERNIER RESIDU = 459

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	7	1.53	1029.49	2.21
2	LEU	L	56	12.20	6332.48	13.57
3	ILE	I	21	4.58	2374.68	5.09
4	MET	M	8	1.74	1048.32	2.25
5	VAL	V	31	6.75	3071.17	6.58
6	SER	S	32	6.97	2784.96	5.97
7	PRO	P	26	5.66	2523.30	5.41
8	THR	T	27	5.88	2728.35	5.85
9	ALA	A	78	16.99	5541.12	11.88
10	TYR	Y	3	0.65	489.18	1.05
11	*	*	0	0.00	0.00	0.00
12	HIS	H	5	1.09	685.30	1.47
13	GLN	Q	13	2.83	1664.78	3.57
14	ASN	N	10	2.18	1140.40	2.44
15	LYS	K	10	2.18	1280.90	2.75
16	ASP	D	19	4.14	2185.57	4.68
17	GLU	E	24	5.23	3096.96	6.64
18	CYS	C	10	2.18	1030.10	2.21
19	TRP	W	2	0.44	372.16	0.80
20	ARG	R	29	6.32	4526.90	9.70
21	GLY	G	48	10.46	2736.96	5.87
22	-	-	0	0.00	0.00	0.00

RESIDUS = 459  
 POIDS MOLECULAIRE = 46661.  
 INDEX DE POLARITE (%) = 37.  
 POINT ISOELECTRIQUE = 6.41  
 DO 260 (1mg/ml) = 0.215 DO 280 (1mg/ml) = 0.315

COBG

DE 1 A 459

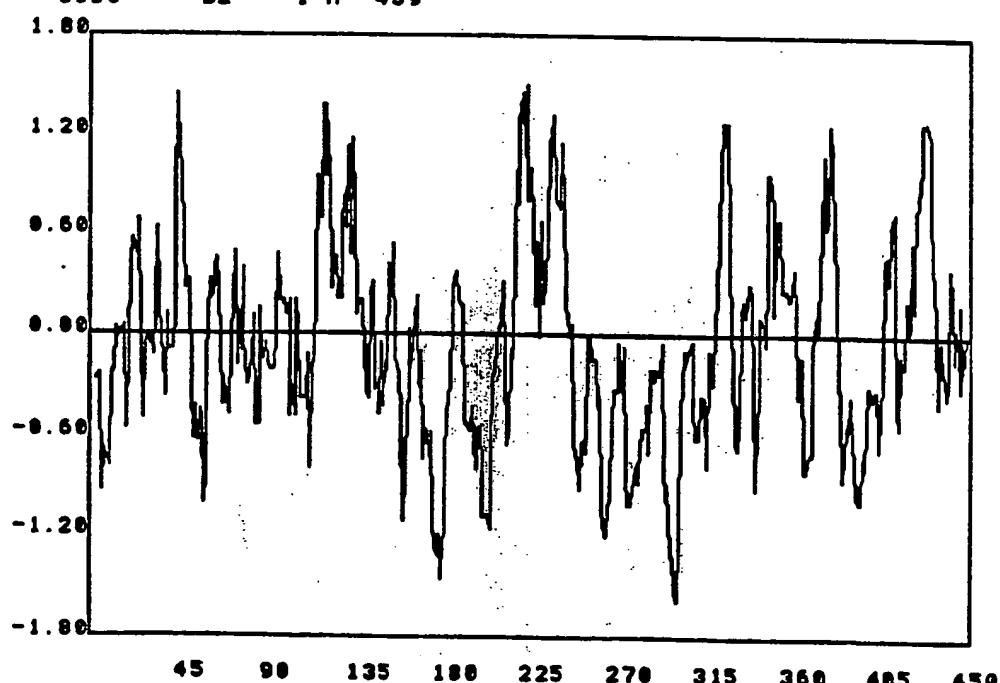


FIGURE 16.4

54/151

GENE cobH ET PROTEINE COBH

SEQUENCE DU FRAGMENT DE 8753 BP, DE 3002 A 3634

MetProGluTyrAspTyrIleArgAspGlyAsnAlaIleTyrGluArgSerPheAlaIleIleArgAlaGluAla  
 ATGCCTGAGTATGATTACATT CGCGATGGCAACGCCATCTACGAGCGTTCCCTCGCCATCATCCCGCGCCGAGGCC  
 3002       3012      3022      3032      3042      3052      3062  
 AspLeuSerArgPheSerGluGluAlaAspLeuAlaValArgMetValHisAlaCysGlySerValGluAla  
 GATCTGTCGCGCTTCTCCGAAGAGGAAGCGGATCTGGCTGTGCGCATGGTGCACGCCTGCGGTTCCGTCGAGGCG  
 3077       3087      3097      3107      3117      3127      3137  
 ThrArgGlnPheValPheSerProAspPheValSerSerAlaArgAlaAlaLeuLysAlaGlyAlaProIleLeu  
 ACCAGGCAGTTCTGTGTTTCTCCGATTCTGTAAGCTCGGCCGTGCGGCGCTGAAAGCCGGTGCGGCCGATCCTC  
 3152       3162      3172      3182      3192      3202      3212  
 CysAspAlaGluMetValAlaHisGlyValThrArgAlaArgLeuProAlaGlyAsnGluValIleCysThrLeu  
 TCGGATGCCGAGATGGTTGCGCACGGTGTCAACCGCGCCCGTCTGCCGGCCGCAACGAGGTGATCTGCACGCTG  
 3227       3237      3247      3257      3267      3277      3287  
 ArgAspProArgThrProAlaLeuAlaAlaGluIleGlyAsnThrArgSerAlaAlaAlaLeuLysLeuTrpSer  
 CGCGATCCTCGCACGCCCGCACTTGC GGCGAGATCGGCAACACCCGCTCCGCCAGCCCTGAAGCTCTGGAGC  
 3302       3312      3322      3332      3342      3352      3362  
 GluArgLeuAlaGlySerValValAlaIleGlyAsnAlaProThrAlaLeuPhePheLeuLeuGluMetLeuArg  
 GAGCGGCTGGCCGGTTCGGTGGTCGCGATCGGCAACCGCGCCGACGGCGTGGTCTTCCTCTGGAAATGCTGCC  
 3377       3387      3397      3407      3417      3427      3437  
 AspGlyAlaProLysProAlaAlaIleLeuGlyMetProValGlyPheValGlyAlaAlaGluSerLysAspAla  
 GACGGCGCGCCGAAGCCGGCGCAATCCCTCGGCATGCCCGTCGGTTTCGTCGGTGC GGCGGAATCGAAGGATGCG  
 3452       3462      3472      3482      3492      3502      3512  
 LeuAlaGluAsnSerTyrGlyValProPheAlaIleValArgGlyArgLeuGlyGlySerAlaMetThrAlaAla  
 CTGGCCGAGAACCTCTATGGCGTTCCCTCGCCATCGTGC CGGCCCTCGGC GGAGTGCCATGACGGCGGCA  
 3527       3537      3547      3557      3567      3577      3587  
 AlaLeuAsnSerLeuAlaArgProGlyLeu\*\*\*  
 GCGCTTAACTCGCTCGAGGCCGGCGCTGTGA  
 3602       3612      3622      3632      3632

FIGURE 16.5

55/151

## PROTEINE COBH

PREMIER RESIDU = 1  
 DERNIER RESIDU = 210

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	9	4.29	1323.63	6.00
2	LEU	L	20	9.52	2261.60	10.26
3	ILE	I	10	4.76	1130.80	5.13
4	MET	M	6	2.86	786.24	3.57
5	VAL	V	14	6.67	1386.98	6.29
6	SER	S	14	6.67	1218.42	5.53
7	PRO	P	12	5.71	1164.60	5.28
8	THR	T	7	3.33	707.35	3.21
9	ALA	A	40	19.05	2841.60	12.89
10	TYR	Y	4	1.90	652.24	2.96
11	*	*	0	0.00	0.00	0.00
12	HIS	H	2	0.95	274.12	1.24
13	GLN	Q	1	0.48	128.06	0.58
14	ASN	N	6	2.86	684.24	3.10
15	LYS	K	4	1.90	512.36	2.32
16	ASP	D	9	4.29	1035.27	4.70
17	GLU	E	14	6.67	1806.56	8.19
18	CYS	C	3	1.43	309.03	1.40
19	TRP	W	1	0.48	186.08	0.84
20	ARG	R	17	8.10	2653.70	12.03
21	GLY	G	17	8.10	969.34	4.40
22	-	-	0	0.00	0.00	0.00

RESIDUS = 210

POIDS MOLECULAIRE = 22050.

INDEX DE POLARITE (%) = 35.

POINT ISOELECTRIQUE = 6.22

DO 260 (1mg/ml) = 0.291 DO 280 (1mg/ml) = 0.467

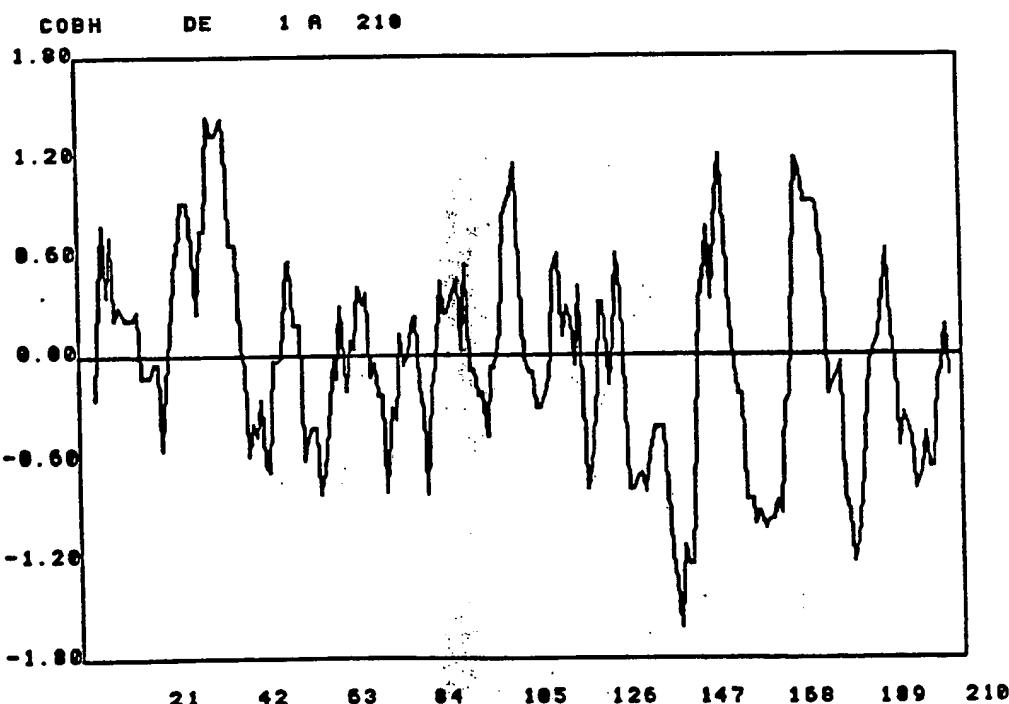


FIGURE 16.6

56 / 151

## GENE COBI ET PROTEINE COBI

SEQUENCE DU FRAGMENT DE 8753 BP, DE 3631 A 4368

MetSerGlyValGlyValGlyArgLeuIleGlyValGlyThrGlyProGlyAspProGluLeuLeuThrValLys  
 GTGAGCGGCCGTGGCGCTGATCGGTGTGGGACCGCCCCGGTGAATCCGAACTTTGACGGTCAAG  
 3631      3641      3651      3661      3671      3681      3691  
 AlaValLysAlaLeuGlyGlnAlaAspValLeuAlaTyrPheAlaLysAlaGlyArgSerGlyAsnGlyArgAla  
 GCGGTGAAGGCGCTCGGGCAAGCCGATGTGCTTGCCATTTCGCCAAGGCCGGCGAACGGTAACGGCCGCG  
 3706      3716      3726      3736      3746      3756      3766  
 ValValGluGlyLeuLeuLysProAspLeuValGluLeuProLeuTyrTyrProValThrThrGluIleAspLys  
 GTGGTCGAGGGTCTGCTGAAGCCCGATCTTGTGAGCTGCCGTATACTATCCGGTAGCGACCAGAACATCGACAAG  
 3781      3791      3801      3811      3821      3831      3841  
 AspAspGlyAlaTyrLysThrGlnIleThrAspPheTyrAsnAlaSerAlaGluAlaValAlaAlaHisLeuAla  
 GACGATGGCGCCTACAAGACCCAGATCACCGACTTCTACAATGCCGTGGCCGAAGCGGTAGCGGCACATCTTGCC  
 3856      3866      3876      3886      3896      3906      3916  
 AlaGlyArgThrValAlaValLeuSerGluGlyAspProLeuPheTyrGlySerTyrMetHisLeuHisValArg  
 GCCGGCGCACGGTCGCCGTGCTCAGTGAAGGCGACCCGCTGTCTATGGTCTCATGCACTGCTATGTGCGG  
 3931      3941      3951      3961      3971      3981      3991  
 LeuAlaAsnArgPheProValGluValIleProGlyIleThrAlaMetSerGlyCysTrpSerLeuAlaGlyLeu  
 CTCGCCAATCGTTCCCGGTCGAGGTGATCCCCGGCATTACCGCCATGTCCGGCTGTTGGTCGCTTGCCGGCTG  
 4006      4016      4026      4036      4046      4056      4066  
 ProLeuValGlnGlyAspAspValLeuSerValLeuProGlyThrMetAlaGluAlaGluLeuGlyArgArgLeu  
 CCGCTGGTGCAGGGCGACGACGTGCTCTGGTGCTTCGGGCACCATGGCCGAGGCCGAGCTCGGCCGAGGCTT  
 4081      4091      4101      4111      4121      4131      4141  
 AlaAspThrGluAlaAlaValIleMetLysValGlyArgAsnLeuProIlysIleArgArgAlaLeuAlaAlaSer  
 GCGGATACCGAAGCCGCCGTGATCATGAAGGTGGCGCAATTGCCGAAGATCCGTGGCGCTCGCTGCCTCC  
 4156      4166      4176      4186      4196      4206      4216  
 GlyArgLeuAspGlnAlaValTyrValGluArgGlyThrMetLysAsnAlaAlaMetThrAlaLeuAlaGluLys  
 GGCGTCTCGACCAAGGCCGTCTATGTCGAACCGCGCACGATGAAGAACGCCGAGTACGGCTCTTGGAAAAG  
 4231      4241      4251      4261      4271      4281      4291  
 AlaAspAspGluAlaProTyrPheSerLeuValLeuValProGlyTrpLysAspArgPro\*\*\*  
 GCGGACGACGAGGCCCTATTCTCGCTGGTGCCTCCGGCTGGAAAGGACCGACCATGA  
 4306      4316      4326      4336      4346      4356      4366

FIGURE 16.7

## PROTEINE COBI

57/151

			PREMIER RESIDU = 1
			DERNIER RESIDU = 245
NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	5
2	LEU	L	28
3	ILE	I	7
4	MET	M	7
5	VAL	V	25
6	SER	S	10
7	PRO	P	14
8	THR	T	12
9	ALA	A	34
10	TYR	Y	9
11	*	*	0
12	HIS	H	3
13	GLN	Q	4
14	ASN	N	5
15	LYS	K	11
16	ASP	D	15
17	GLU	E	13
18	CYS	C	1
19	TRP	W	2
20	ARG	R	14
21	GLY	G	26
22	-	-	0
			0.00
			0.00

RESIDUS = 245

POIDS MOLECULAIRE = 25878.

INDEX DE POLARITE (%) = 36.

POINT ISOELECTRIQUE = 6.17

DO 260 (1mg/ml) = 0.512 DO 280 (1mg/ml) = 0.843

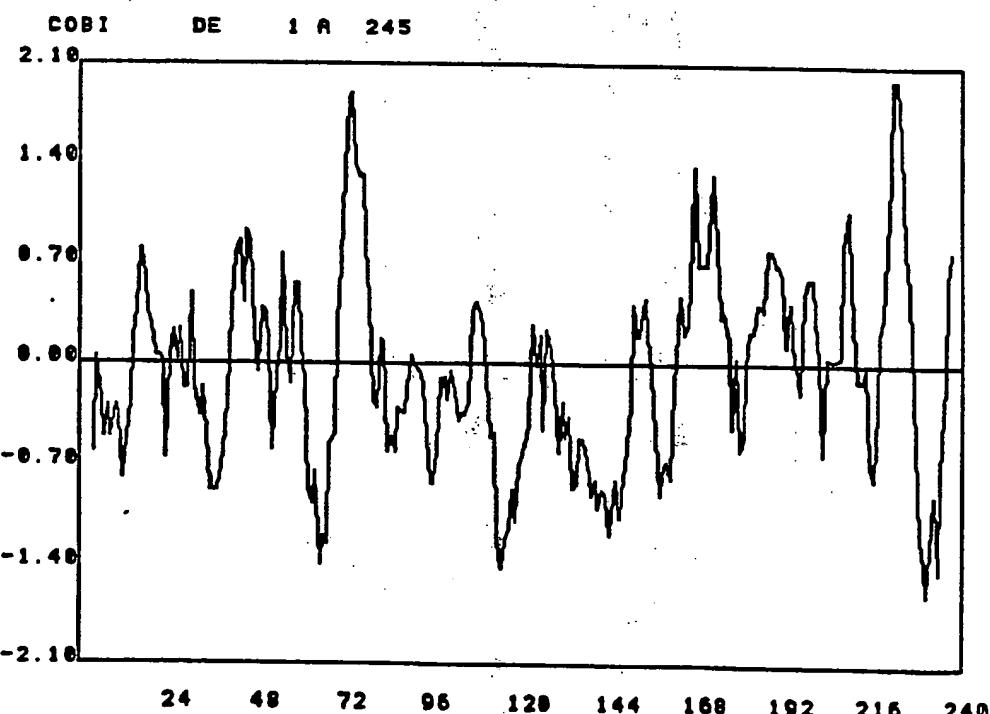


FIGURE 16.8

58 / 151

GENE cobJ ET PROTEINE COBJ

SEQUENCE DU FRAGMENT DE 8753 BP, DE 4365 A 5129

MetThrGlyThrLeuTyrValValGlyThrGlySerAlaLysGlnMetThrProGluThrAlaGluAla  
 ATGACCGGTACGCTCTATGTCGTCGGTACCGGACCGGGCAGCGCCAAGCAGATGACGCCGGAAACGGCGGAAGCC  
 4365 4375 4385 4395 4405 4415 4425  
 ValAlaAlaAlaGlnGluPheTyrGlyTyrPheProTyrLeuAsnLeuArgProAspGlnIleArg  
 GTTGCGGCCGCTCAGGAGTTTACGGCTACTTCCCTATCTGACCGGCTGAACCTCAGACCGGATCAGATCCGT  
 4440 4450 4460 4470 4480 4490 4500  
 ValAlaSerAspAsnArgGluGluLeuAspArgAlaGlnValAlaLeuThrArgAlaAlaAlaGlyValLysVal  
 GTCGCCTCGGACAACCGCGAGGAGCTCGATCGGGCACAGGTCGCGCTGACGCCGGCTGCGAGGCGATCAGAAGGGACCG  
 4515 4525 4535 4545 4555 4565 4575  
 CysMetValSerGlyGlyAspProGlyValPheAlaMetAlaAlaAlaValCysGluAlaIleAspLysGlyPro  
 TGCATGGTCTCCGGTGGCGATCCC GGTTGCTTTGCCATGGCGGCCGCTGCGAGGCGATCAGAAGGGACCG  
 4590 4600 4610 4620 4630 4640 4650  
 AlaGluTrpLysSerValGluLeuValIleThrProGlyValThrAlaMetLeuAlaValAlaAlaArgIleGly  
 GCGGAATGGAAGTCGGTTGAAGTGGTGCATCACGCCGGCGTACCGCGATGCTCGCCGTTGCCGCCGCATCGGC  
 4665 4675 4685 4695 4705 4715 4725  
 AlaProLeuGlyHisAspPheCysAlaIleSerLeuSerAspAsnLeuLysProTrpGluValIleThrArgArg  
 GCGCCGCTCGGTCACTGATTCTGTGCATCTCGCTTCCGACAATCTGAAGCCCTGGGAAGTCATCACCCGGCGT  
 4740 4750 4760 4770 4780 4790 4800  
 LeuArgLeuAlaAlaGluAlaGlyPheValIleAlaLeuTyrAsnProIleSerLysAlaArgProTrpGlnLeu  
 CTCAGGCTGGCGGCCGGAAAGCGGGCTTCGTCATTGCCCTCTACAATCGATCAGCAAGGCGCGGCCCTGGCAGCTC  
 4815 4825 4835 4845 4855 4865 4875  
 GlyGluAlaPheGluLeuLeuArgSerValLeuProAlaSerValProValIlePheGlyArgAlaAlaGlyArg  
 GGTGAGGCCCTCGAGCTCTGCGCAGCGTTCTGCCGGCAAGCGTTCCGGTCATCTCGCCGTTGCCGCCGGCG  
 4890 4900 4910 4920 4930 4940 4950  
 ProAspGluArgIleAlaValMetProLeuGlyGluAlaAspAlaAsnArgAlaAspMetAlaThrCysValIle  
 CCGGACGAAACGGATCGCGGTGATGCCGCTCGCGAGGCCGATGCCAACCGCGCCGACATGGCGACCTGCGTCATC  
 4965 4975 4985 4995 5005 5015 5025  
 IleGlySerProGluThrArgIleValGluArgAspGlyGlnProAspLeuValTyrThrProArgPheTyrAla  
 ATCGGCTCGCCGGAGACGCCATCGTCGAGCGCGACGGCCAACCCGATCTCGTCTACACACCGCGCTTCTATGCA  
 5040 5050 5060 5070 5080 5090 5100  
 GlyAlaSerGln\*\*\*  
 GGGGCCGAGCCAGTGA  
 5115 5125

FIGURE 16.9

PROTEINE COBJ

59/151

PREMIER RESIDU = 1  
DERNIER RESIDU = 254

	NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	8	3.15
2	LEU	L	20	7.87
3	ILE	I	13	5.12
4	MET	M	7	2.76
5	VAL	V	23	9.06
6	SER	S	11	4.33
7	PRO	P	18	7.09
8	THR	T	12	4.72
9	ALA	A	40	15.75
10	TYR	Y	7	2.76
11	*	*	0	0.00
12	HIS	H	1	0.39
13	GLN	Q	7	2.76
14	ASN	N	5	1.97
15	LYS	K	6	2.36
16	ASP	D	13	5.12
17	GLU	E	16	6.30
18	CYS	C	4	1.57
19	TRP	W	3	1.18
20	ARG	R	19	7.48
21	GLY	G	21	8.27
22	-	-	0	0.00

RESIDUS = 254  
 POIDS MOLECULAIRE = 27088.  
 INDEX DE POLARITE (%) = 35.  
 POINT ISOELECTRIQUE = 5.43  
 DO 260 (1mg/ml) = 0.575 DO 280 (1mg/ml) = 0.922

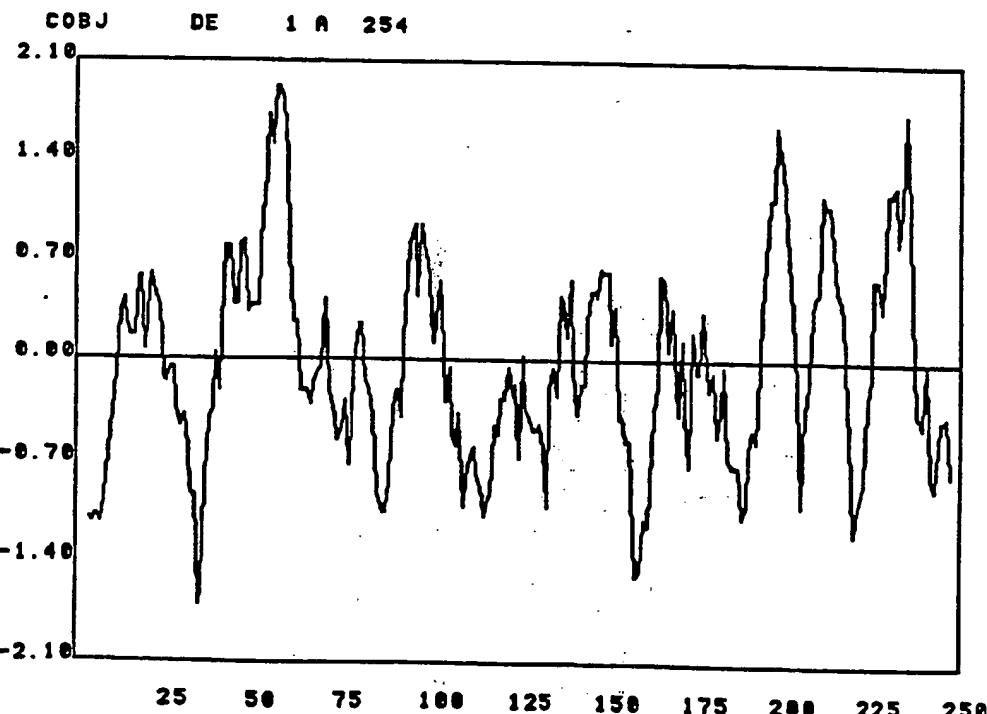


FIGURE 16.10

GENE cobK ET PROTEINE COBK

60/151

SEQUENCE DU FRAGMENT ECORI-ECORI DE 8753 PB DE 2861 A 3646 SUR LE BRIN  
COMPLEMENTAIRE

M	A	G	S	L	F	D	T	S	A	M	E	K	P	R	I	L	I	L	G	G	T	T	E	A	
ATGGCGGGTTTCGCTGTCGACACGTCAAGCCATTGGAAAAACCTCGTATTCTGATTCTGGTGGCACCAACCGAGGCA																									
2861	2871	2881	2891	2901	2911	2921	2931																		
R	E	L	A	R	R	L	A	E	D	V	R	Y	D	T	A	I	S	L	A	G	R	T	A	D	
CGCGAACTCGCGGCCGCTTGGCCGAAGATGTCGCCTACGACACCGCCATCTCGCTGGCGGCCGACCCGGAC																									
2936	2946	2956	2966	2976	2986	2996	3006																		
P	R	P	Q	P	V	K	T	R	I	G	G	F	G	G	A	D	G	L	A	H	F	V	H	D	
CCGC CGCC CGCAGCCGGTCAAGACGCCATCGCGGGCTTGGCGGCCGATGGCTGGCGATTCGTGCGATGAC																									
3011	3021	3031	3041	3051	3061	3071	3081																		
E	N	I	A	L	L	V	D	A	T	H	P	F	A	A	R	I	S	H	N	A	A	D	A	A	
GAAAACATCGCGCTGCTGGTCGATGCGACGCCCTTGCCGCACGCATTCGCACAACGCCGGACCGAGCG																									
3086	3096	3106	3116	3126	3136	3146	3156																		
Q	R	T	G	V	A	L	I	A	L	R	R	P	E	W	V	P	L	P	G	D	R	W	T	A	
CAAAGAACCGCGTTCGCTTATGCCCTCCGCCGACCGGAATGGGTGCCCTGCCTGGCGACCGCTGGACTGCT																									
3161	3171	3181	3191	3201	3211	3221	3231																		
V	D	S	V	V	E	A	V	S	A	L	G	D	R	R	R	R	V	F	L	A	I	G	R	Q	
GTCGATAGCGTTGTCGAGGCCGTCAGCGCGCTCGTGCATGGCGACGCCGCGCTTCCTGGCGATAGGTGCGACAG																									
3236	3246	3256	3266	3276	3286	3296	3306																		
E	A	F	H	F	E	V	A	P	Q	H	S	Y	V	I	R	S	V	D	P	V	T	P	P	L	
GAAGCTTCCACTTCGAGGTGCGCCGCAGCACAGCTACGGCATCCGCAGCGTGCATCCGGTGCACGCCGCCGCTT																									
3311	3321	3331	3341	3351	3361	3371	3381																		
N	L	P	D	Q	E	A	I	L	A	T	G	P	F	A	E	A	D	E	A	A	L	L	R	S	
AATCTGCCCGACCAGGAGGCGATCTGGCGACCGGTCCCTTGCGGAAGCCGACGAAGCCGCGTTGCTCAGGAGT																									
3386	3396	3406	3416	3426	3436	3446	3456																		
R	Q	I	D	V	I	V	A	K	N	S	G	G	S	A	T	Y	G	K	I	A	A	A	R	R	
CGGCAGATCGATGTGATCGTCGCCAAGAACAGCGGTGGCAGCGCCACCTACGGCAAGATTGCCGCAGCGCGCCGG																									
3461	3471	3481	3491	3501	3511	3521	3531																		
L	G	I	E	V	I	M	V	E	R	R	K	P	A	D	V	P	T	V	G	S	C	D	E	A	
CTCGGCATCGAGGTGATCATGGTCAGCGGCCAAGCCCGCGACGTGCGACGGTCGGCAGTTGCGACGAGGCA																									
3536	3546	3556	3566	3576	3586	3596	3606																		
L	N	R	I	A	H	W	L	A	P	A															
CTCAACCGCATCGCTCACTGGCTGCCCTGCATGA																									
3611	3621	3631	3641																						

FIGURE 16-11

61/151

NOM = COBK

PREMIER RESIDU = 1  
DERNIER RESIDU = 261

	NOMBRE	% NOMB	POIDS	% POIDS
1. PHE	F	8	3.07	1176.55
2. LEU	L	22	8.43	2487.85
3. ILE	I	16	6.13	1809.34
4. MET	M	3	1.15	393.12
5. VAL	V	21	8.05	2080.44
6. SER	S	12	4.60	1044.38
7. PRO	P	17	6.51	1649.90
8. THR	T	13	4.98	1313.62
9. ALA	A	42	16.09	2983.56
10. TYR	Y	3	1.15	489.19
11. *	*	0	0.00	0.00
12. HIS	H	7	2.68	959.41
13. GLN	Q	6	2.30	768.35
14. ASN	N	5	1.92	570.21
15. LYS	K	5	1.92	640.47
16. ASP	D	17	6.51	1955.46
17. GLU	E	15	5.75	1935.64
18. CYS	C	1	0.38	103.01
19. TRP	W	3	1.15	558.24
20. ARG	R	26	9.96	4058.63
21. GLY	G	19	7.28	1083.41
22.	-	0	0.00	0.00

RESIDUS = 261

Masse moleculaire (monoisotopique) = 28078.7988

Masse moleculaire (moyenne) = 28096.0195

INDEX DE POLARITE (%) = 40.61

POINT ISOELECTRIQUE = 7.54

DO 260 (1mg/ml) = 0.509 DO 280 (1mg/ml) = 0.721

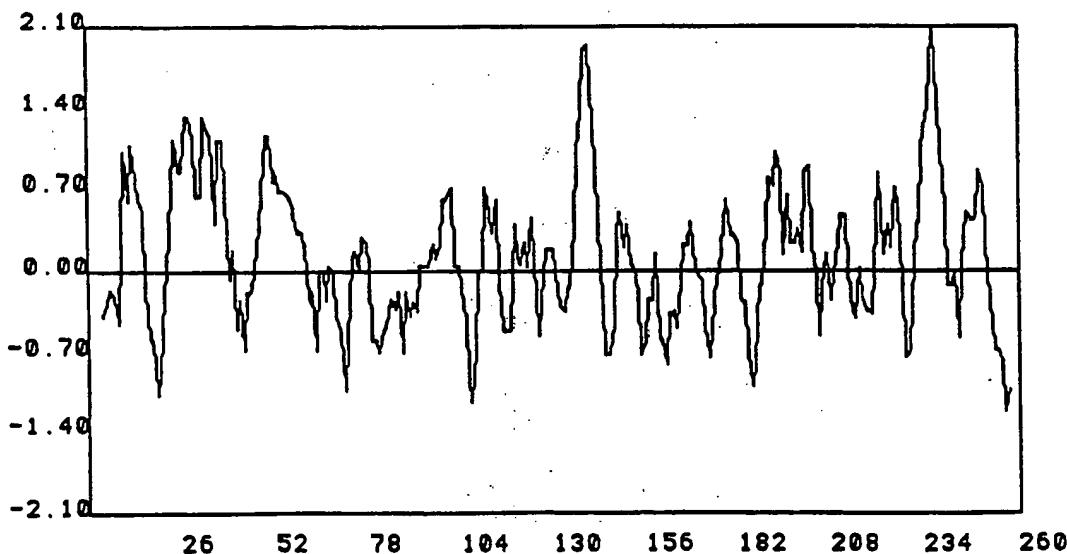
PROFIL D'HYDROPHILICITE DE LA PROTEINE  
COBK DE 1 A 261

FIGURE 16.12

62/151

GENE cobl ET PROTEINE COBL  
SEQUENCE DU FRAGMENT DE 8753 BP, DE 5862 A 7103

MetAlaAspValSerAsnSerGluProAlaIleValSerProTrpLeuThrValIleGlyIleGlyGluAspGly  
ATGGCTGACGTGTCGAACAGCGAACCCGCCATAGCTCCCCCTGGCTGACCGTCATCGTATCGGTGAGGATGGT  
5862 5872 5882 5892 5902 5912 5922  
ValAlaGlyLeuGlyAspGluAlaLysArgLeuIleAlaGluAlaProValValTyrGlyGlyHisArgHisLeu  
GTAGCGGGTCTCGGCCACGGCCAAGCGGCTGATCGCGAACGCACGGTCTACGGCGGCCATCGTCATCTG  
5937 5947 5957 5967 5977 5987 5997  
GluLeuAlaAlaSerLeuIleThrGlyGluAlaHisAsnTrpLeuSerProLeuGluArgSerValValGluIle  
GAGCTCGCCGCCCTCCCATCACCGCGAACGCACAAATTGGCTAAGCCCCCTCGAACGCTCGGTGAGATC  
6012 6022 6032 6042 6052 6062 6072  
ValAlaArgArgGlySerProValValLeuAlaSerGlyAspProPhePheGlyValGlyValThrLeu  
GTCGCCGTCGCCAGCCCCGGTGGTGCTTGCTCGGGCACCGTTCTTCGGCGTCGGCGTGAACGCTG  
6087 6097 6107 6117 6127 6137 6147  
AlaArgArgIleAlaSerAlaGluIleArgThrLeuProAlaProSerSerIleSerLeuAlaAlaSerArgLeu  
GCGCGCCGCATCGCCTCGGCCGAAATACGCACGCTTCCGGCGCGTCGATCAGTCTTGCCGCCTCGCGCCTC  
6162 6172 6182 6192 6202 6212 6222  
GlyTrpAlaLeuGlnAspAlaThrLeuValSerValHisGlyArgProLeuAspLeuValArgProHisLeuHis  
GGCTGGCGCTGCAGGATGCGACGCTCGTCCGTACATGGGCGGCCGCTGGATCTGGTGCGACCGCATTGCA  
6237 6247 6257 6267 6277 6287 6297  
ProGlyAlaArgValLeuThrLeuThrSerAspGlyAlaGlyProArgAspLeuAlaGluLeuLeuValSerSer  
CCGGGGCGCGTGTGCTTACGCTCACGTCGGACGGTGCAGGTCGGTCCGCGAGACCTTGCCGAGCTTCTGGTTTCAAGC  
6312 6322 6332 6342 6352 6362 6372  
GlyPheGlyGlnSerArgLeuThrValLeuGluAlaLeuGlyAlaGlyGluArgValThrThrGlnIleAla  
GGCTTCGGTCAGTCGGACTGACCGTGCTCGAACCGCTGGCGCCGGCGAACGGGTGACGACGACAGATCGCC  
6387 6397 6407 6417 6427 6437 6447  
AlaArgPheMetLeuGlyLeuValHisProLeuAsnValCysAlaIleGluValAlaAlaAspGluGlyAlaArg  
GCGCGCTTCATGCTCGGCCCTCGCATCTTGAAACGTCTGCGCCATTGAGGTGGCGGCCGACGAGGGCGCGC  
6462 6472 6482 6492 6502 6512 6522  
IleLeuProLeuAlaAlaGlyArgAspAspAlaLeuPheGluHisAspGlyGlnIleThrLysArgGluValArg  
ATCCTGCCGCTTGCCGCCGCGACGATGCGCTGTCGAACATGACGGCAGATCACCAAGCGCGAGGTGCGG  
6537 6547 6557 6567 6577 6587 6597  
AlaLeuThrLeuSerAlaLeuAlaProArgLysGlyGluLeuLeuTrpAspIleGlyGlySerGlySerIle  
GCGCTGACGCTGTCGGCACTCGCACCGCGCAAGGGCGAACTGCTATGGGACATCGCGCGCTCCGGCTCGATC  
6612 6622 6632 6642 6652 6662 6672  
GlyIleGluTrpMetLeuAlaAspProThrMetGlnAlaIleThrIleGluValGluProGluArgAlaAlaArg  
GGCATCGAACGCGATCCGACCATGCAGGGCGATCACCATCGAGGGTTGAGCCGGAGCGGGCAGCGCGC  
6687 6697 6707 6717 6727 6737 6747  
IleGlyArgAsnAlaThrMetPheGlyValProGlyLeuThrValValGluGlyGluAlaProAlaAlaLeuAla  
ATCGGCCAACCGCGACGATGTTCGCGTGGCGCCGGCTGACGGITGTCGAAGGGCGAGGCAGGCCGGCGCTTGCC  
6762 6772 6782 6792 6802 6812 6822  
GlyLeuProGlnProAspAlaIlePheIleGlyGlyGlySerGluAspGlyValMetGluAlaAlaIleGlu  
GGCTGCCAACCGGACCGCATCTCATCGCGCGGGCAGCGAACGCGCTCATGGAAGCAGCGATCGAG  
6837 6847 6857 6867 6877 6887 6897  
AlaLeuLysSerGlyGlyArgLeuValAlaAsnAlaValThrAspMetGluAlaValLeuLeuAspHisHis  
GCGCTCAAGTCAGGCGGACGGCTGGTTGCCAACCGGGTGACGACGGACATGGAAGCGGTGCTGCTCGATCATCAC  
6912 6922 6932 6942 6952 6962 6972  
AlaArgLeuGlyGlySerLeuIleArgIleAspIleAlaArgAlaGlyProIleGlyGlyMetThrGlyTrpLys  
GCGCGGCTCGGCCGGTCTGCTGATCCGCATCGATATCGCGGTGAGGACCCATCGCGGCATGACCGGCTGGAAG  
6987 6997 7007 7017 7027 7037 7047  
ProAlaMetProValThrGlnTrpSerTrpThrLysGly\*\*\*  
CCGGCCATGCCGGTCACCAATGGTCGGACGAAGGGCTAA  
7062 7072 7082 7092 7102

FIGURE 16.13

PROTEINE COBL

63/151

PREMIER RESIDU = 1  
DERNIER RESIDU = 413

	NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	8	1.94
2	LEU	L	47	11.38
3	ILE	I	26	6.30
4	MET	M	9	2.18
5	VAL	V	34	8.23
6	SER	S	25	6.05
7	PRO	P	24	5.81
8	THR	T	21	5.08
9	ALA	A	56	13.56
10	TYR	Y	1	0.24
11	*	*	0	0.00
12	HIS	H	10	2.42
13	GLN	Q	7	1.69
14	ASN	N	5	1.21
15	LYS	K	6	1.45
16	ASP	D	19	4.60
17	GLU	E	27	6.54
18	CYS	C	1	0.24
19	TRP	W	8	1.94
20	ARG	R	28	6.78
21	GLY	G	51	12.35
22	-	-	0	0.00

RESIDUS = 413  
 POIDS MOLECULAIRE = 42911.  
 INDEX DE POLARITE (%) = 36.  
 POINT ISOELECTRIQUE = 5.70  
 DO 260 (1mg/ml) = 0.754 DO 280 (1mg/ml) = 1.064

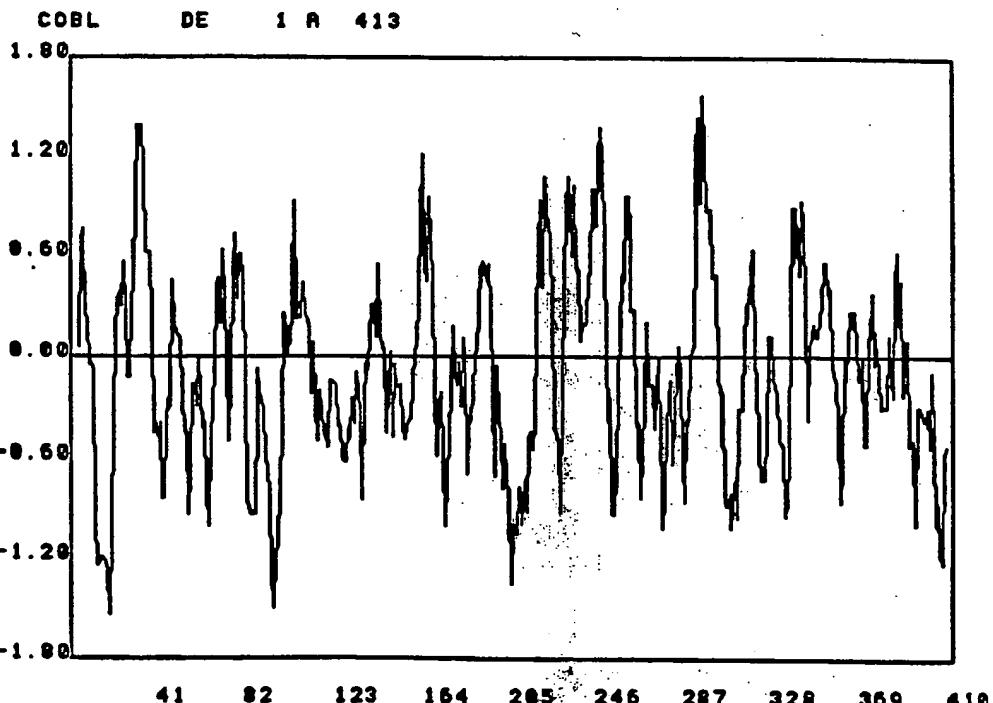


FIGURE 16.14

64 / 151

GENE cobM ET PROTEINE COBM

SEQUENCE DU FRAGMENT DE 8753 BP, DE 7172 A 7930

MetThrValHisPheIleGlyAlaGlyProGlyAlaAlaAspLeuIleThrValArgGlyArgAspLeuIleGly  
 ATGACGGTACATTTCATCGGCCGCCGGCCGGAGCCGAGACCTGATCACGGTGCCTGGTCGCGACCTGATCGGG  
 7172 7182 7192 7202 7212 7222 7232  
 ArgCysProValCysLeuTyrAlaGlySerIleValSerProGluLeuLeuArgTyrCysProProGlyAlaArg  
 CGCTGCCGGTCTGCCTTACGCCGGCTCGATCGTCTCGCCGGAGCTGTCGATATTGCCCGCCGGCGCCCCGC  
 7247 7257 7267 7277 7287 7297 7307  
 IleValAspThrAlaProMetSerLeuAspGluIleGluAlaGluTyrValLysAlaGluAlaGluGlyLeuAsp  
 ATTGTGATACGGCGCCGATGTCCCTCGACGAGATCGAGGCGGAGTATGTGAAGGCCGAAGCCGAAGGGCTCGAC  
 7322 7332 7342 7352 7362 7372 7382  
 ValAlaArgLeuHisSerGlyAspLeuSerValTrpSerAlaValAlaGlnIleArgArgLeuGluLysHis  
 GTGGCGCGGCTTCATTGGCGACCTTCGGTCTGGAGTGCTGTGGCCGACAGATCCCGGGCTCGAGAACGAT  
 7397 7407 7417 7427 7437 7447 7457  
 GlyIleAlaTyrThrMetThrProGlyValProSerPheAlaAlaAlaSerAlaLeuGlyArgGluLeuThr  
 GGCATCGCCTATACGATGACGCCGGCGTTCCCTTGCGCGGCGCTTCAGCGCTCGTCGCGAATTGACC  
 7472 7482 7492 7502 7512 7522 7532  
 IleProAlaValAlaGlnSerLeuValLeuThrArgValSerGlyArgAlaSerProMetProAsnSerGluThr  
 ATTCCGGCCGTGGCCCAGAGCCTGGTGCTGACCCGGCTTCGGGCCGCGCTCGCCGATGCCGAACTCAGAAACG  
 7547 7557 7567 7577 7587 7597 7607  
 LeuSerAlaPheGlyAlaThrGlySerThrLeuAlaIleHisLeuAlaIleHisAlaLeuGlnGlnValValGlu  
 CTTTCGCTTCGGCGCTACGGGATCGACGCTGGCAATCCACCTTGCGATCCATGCCGTTAGCAGGTGGTCGAG  
 7622 7632 7642 7652 7662 7672 7682  
 GluLeuThrProLeuTyrGlyAlaAspCysProValAlaIleValValLysAlaSerTrpProAspGluArgVal  
 GAACTGACGCCGCTACGGTGCCGACTGCCCGGTCGCTCAAGGCCCTGGCCGGACGAAACGCGTG  
 7697 7707 7717 7727 7737 7747 7757  
 ValArgGlyThrLeuGlyAspIleAlaAlaLysValAlaGluGluProIleGluArgThrAlaLeuIlePheVal  
 GTGCGCGGCACGCTCGGTGACATGCCGCAAGGTGGCGGAAGAGCCGATCGAGCGCACGGCGCTGATCTCGTC  
 7772 7782 7792 7802 7812 7822 7832  
 GlyProGlyLeuGluAlaSerAspPheArgGluSerSerLeuTyrAspProAlaTyrGlnArgArgPheArgGly  
 GGTCCGGGGCTCGAACGCTCCGATTCCGTGAAAGCTCGCTACGATCCCGCTATCAGCGCGCTTCAGAGGG  
 7847 7857 7867 7877 7887 7897 7907  
 ArgGlyGlu  
 CGCGGCGAA  
 7922 7932 7942 7952 7962 7972 7982

FIGURE 16.15

PROTEINE COBM

65/451

PREMIER RESIDU = 1  
DERNIER RESIDU = 253

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	6	2.37	882.42	3.29
2	LEU	L	24	9.49	2713.92	10.11
3	ILE	I	15	5.93	1696.20	6.32
4	MET	M	4	1.58	524.16	1.95
5	VAL	V	22	8.70	2179.54	8.12
6	SER	S	18	7.11	1566.54	5.84
7	PRO	P	17	6.72	1649.85	6.15
8	THR	T	13	5.14	1313.65	4.89
9	ALA	A	35	13.83	2486.40	9.26
10	TYR	Y	7	2.77	1141.42	4.25
11	*	*	0	0.00	0.00	0.00
12	HIS	H	5	1.98	685.30	2.55
13	GLN	Q	5	1.98	640.30	2.39
14	ASN	N	1	0.40	114.04	0.42
15	LYS	K	4	1.58	512.36	1.91
16	ASP	D	11	4.35	1265.33	4.71
17	GLU	E	19	7.51	2451.76	9.13
18	CYS	C	4	1.58	412.04	1.53
19	TRP	W	2	0.79	372.16	1.39
20	ARG	R	19	7.51	2965.90	11.05
21	GLY	G	22	8.70	1254.44	4.67
22	-	-	0	0.00	0.00	0.00

RESIDUS = 253

POIDS MOLECULAIRE = 26846.

INDEX DE POLARITE (%) = 38.

POINT ISOELECTRIQUE = 5.58

DO 260 (1mg/ml) = 0.461 DO 280 (1mg/ml) = 0.724

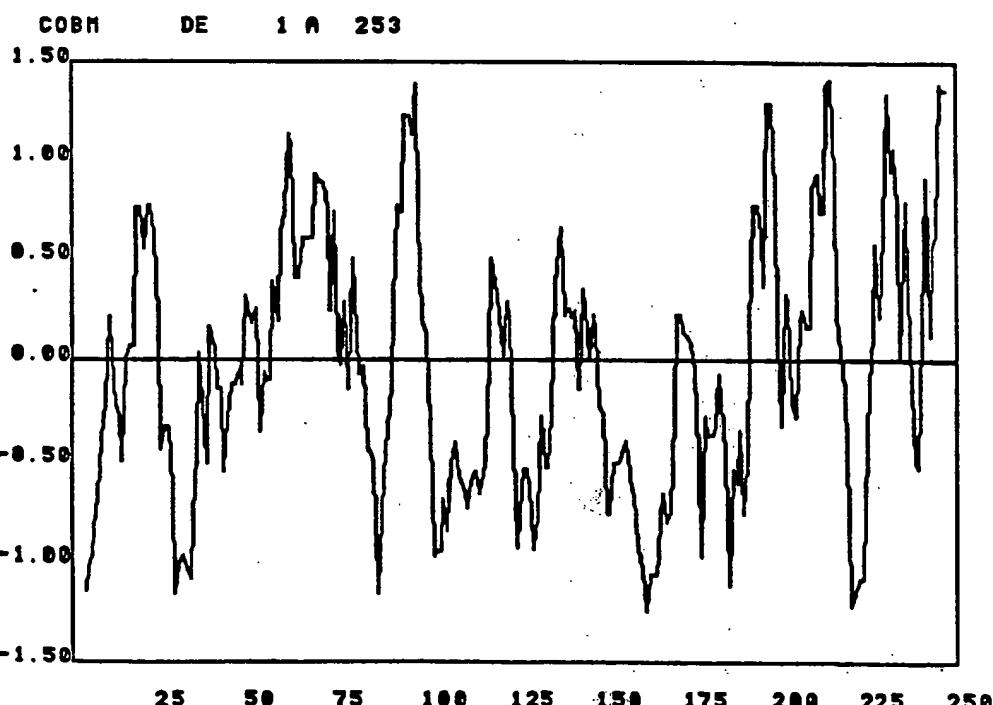


FIGURE 16.16

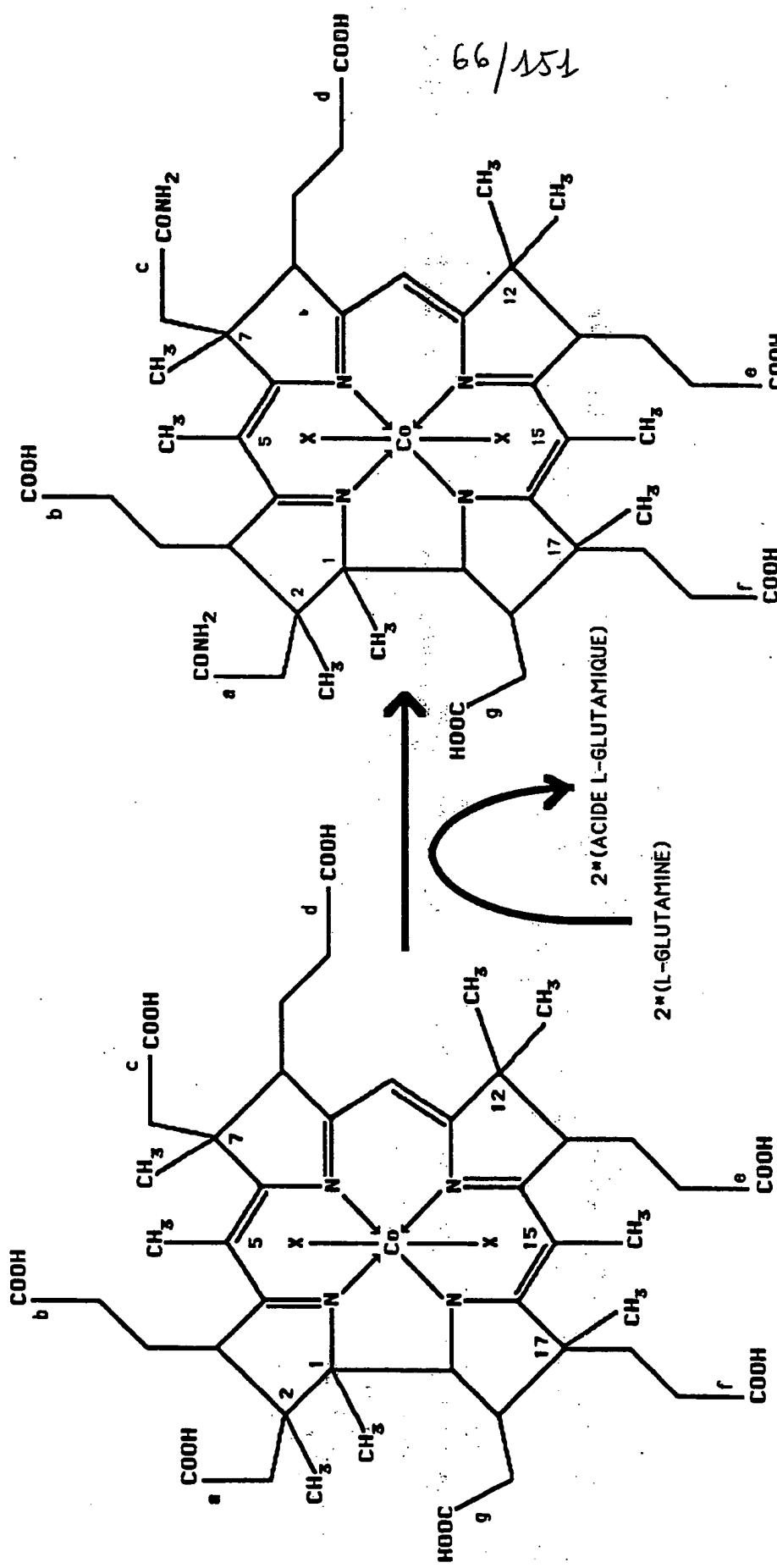


FIGURE 17

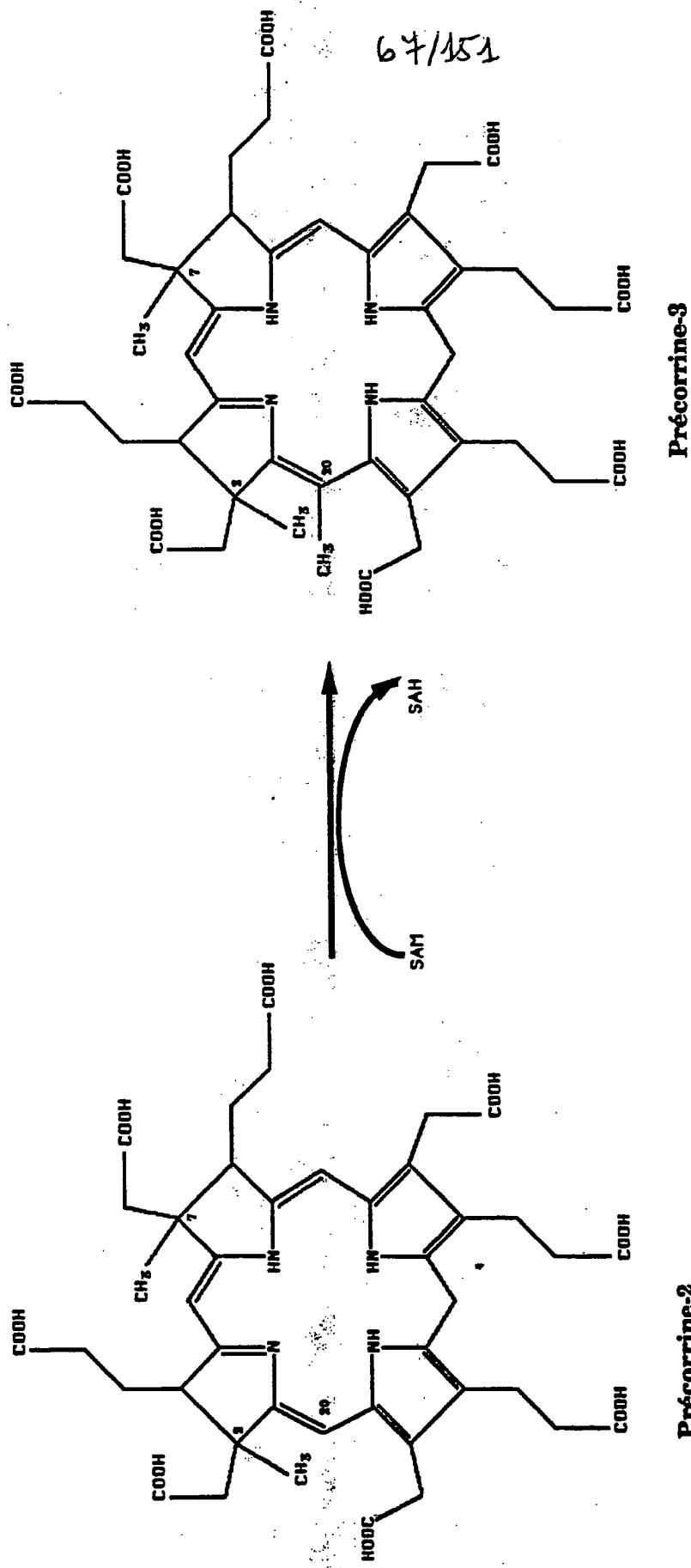


FIGURE 18

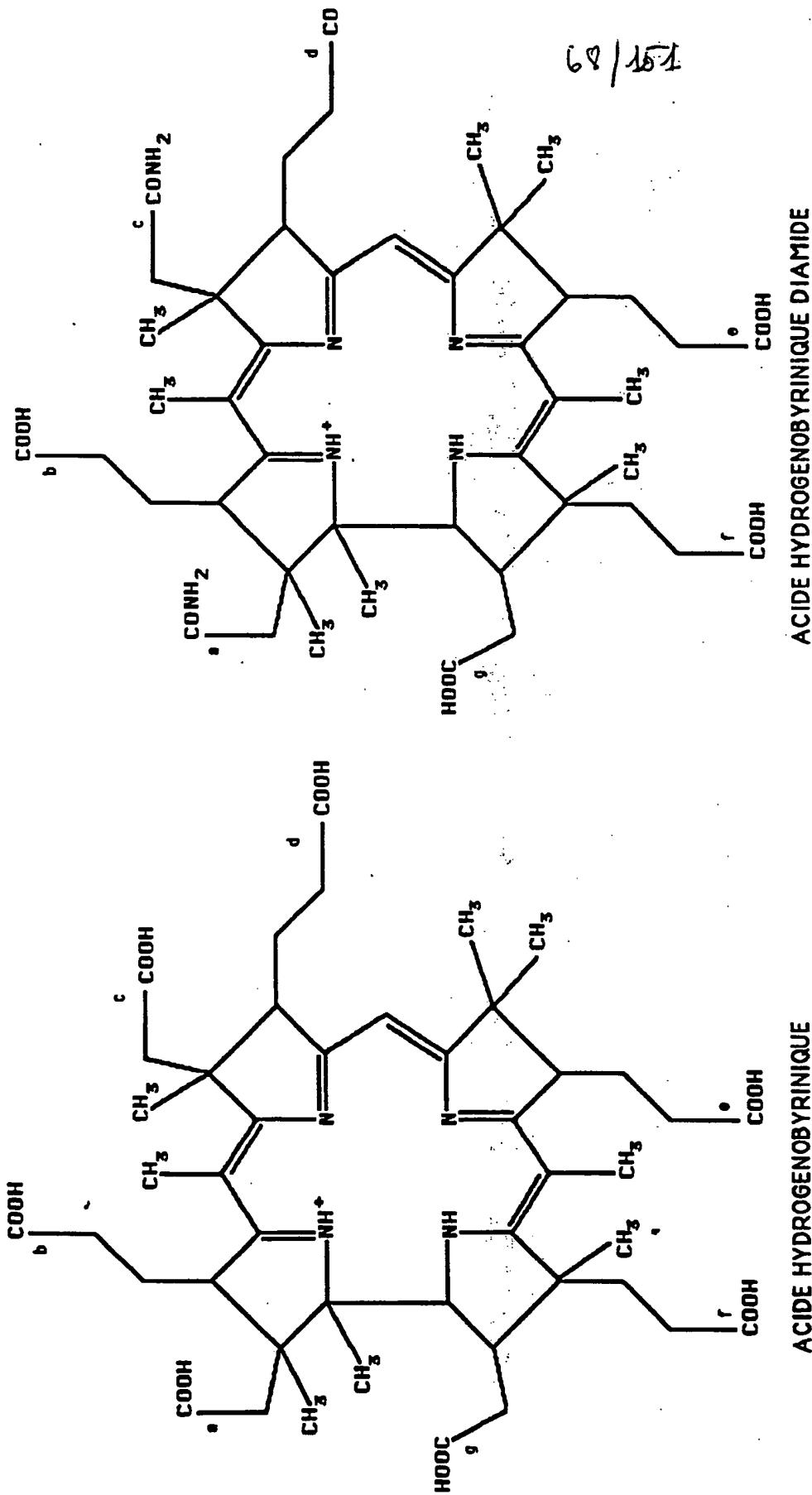


FIGURE 19

69/151

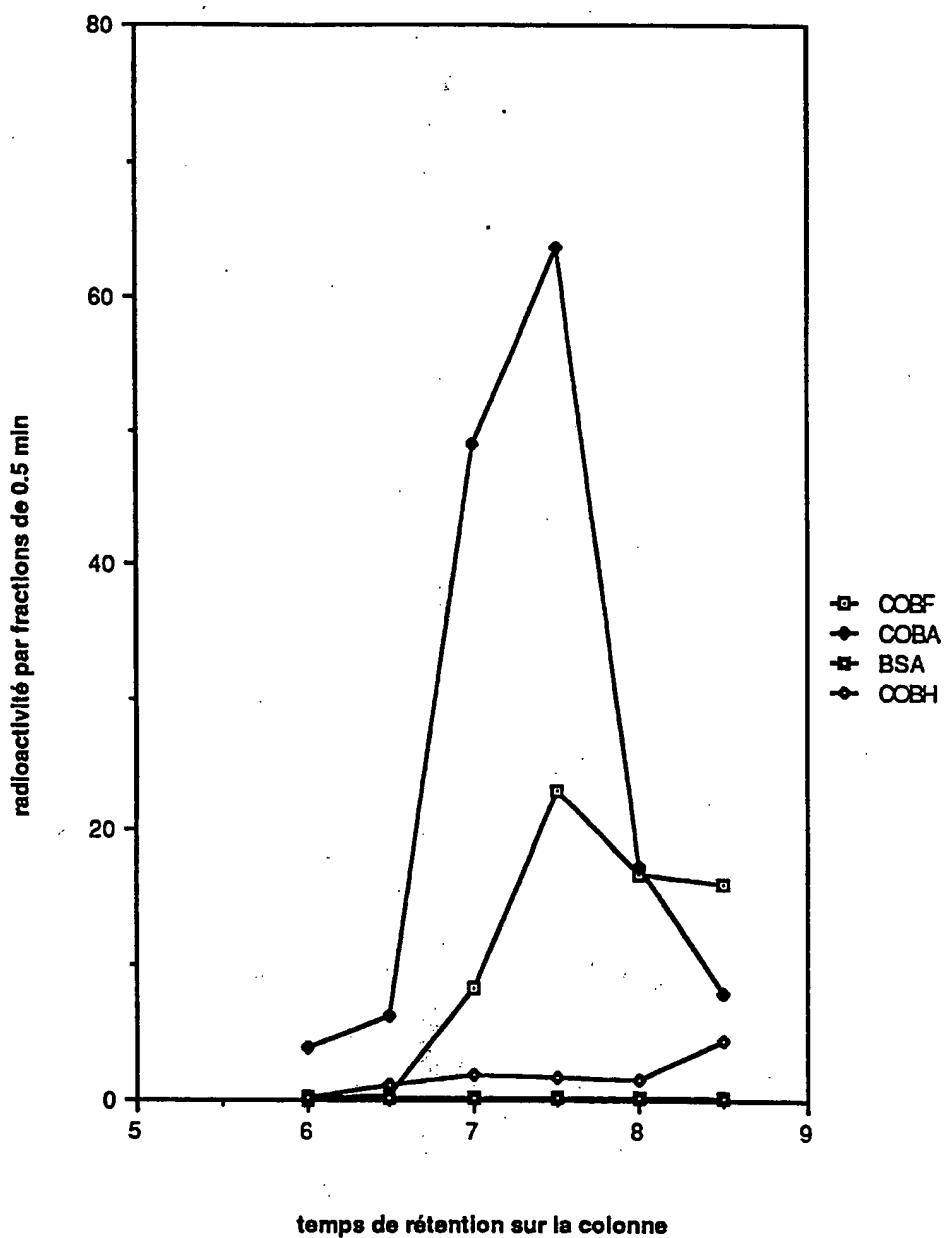


FIGURE 20

70 / 151

	10	20	30	40
<b>COBA</b>	<u>LPALLEKGSVWL-VGAGPGDPGLLTLHAANALRQADVIVH</u>			
	- - -	- - -	- - -	- - -
<b>COBI</b>	VSGVGVGRL-IGVGTGPGDPELLTVKAVKALGQADVLAY			
	10	20	30	40

## région 1

90	100	110	120	130	140	150
COBA	LARAGNRVLRLKGDPFVFRGGEEALTLVEHQVFIRVPGITAGIGGLAYAGIPVTHREVNHAVTFLTG	====	====	====	====	====
COBI	LA-AGRIVAVLSEGDPFLFYGSYMHHLVRLANR-FPVEVIPHITAMSGCWSLAGLPLVQ-G-DDVLSVLPG	110	120	130	140	150
						160

région 2

### région 3

FIGURE 21

71 /151

ALIGNEMENT ENTRE LES PROTEINES COBA DE PSEUDOMONAS DENITRIFICANS ET CYSG DE  
ESCHERICHIA COLI

SEQUENCE COBA  
SEQUENCE CYSG

10	20	30	40	50	60	70
DDLFLAGLPALEKGSVWLGVAGPGDPGLLTLHAANALRQADVIVHDALVNED				CLKLARP	GAVLEFAGKRGG	
EOLINE-PLDHRGEVVVLGVAGPGDAGLTLKGLOOIQOADVYYDRLVSDI				IMNLVRRDADRVFVGKRG		
210	220	230	240	250	260	270
80	90	100	2 110	120	130	140
KPSPKORDISLRVELARA	GNRVLRLKGDPFVFGRGGEELTIVEHQPFRIVPGITAGIGGLAYAGIP					
YHCVPQEEINQILLREAQK	GKRVVRLKGDPFFGRGGEELTICNAGIPFSVPPGITAASGCSAYSGIP					
280	290	300	310	320	330	340
150	160	170	180	190	200	210
VTHREVNHAVTFLTGHDSGLVPDRINWQGIASGSPVIVMMAMKHIGAITANLIAGGRSPDEPVAFVCN						
LTHRDIQSVRLLTGHIKTG-G-E-LDWENLAAEKQTLVFMGLNQAATIQQKLIEHGMGPGEMPVAIEN						
350	360	370	380	390	400	
220	230	240	250			
AATPQQAVLETTLARAEADVAAGLEPPAIVVVGEVVRLRAALDWIG						
GTAVTQRVIDGTLTQL-GELAQO-MNSPSLIIIGRVVGLRDKLNWFS						
420	430	440	450			

SEQUENCE COBA DE 3 A 259

SEQUENCE CYSG DE 204 A 460

HOMOLOGIE STRICTE EN ACIDE AMINES: 41.6%

FIGURE 22

+2/151

**REGION 1**

CYSG(209-248)	E	P	L	D	H	R	G	E	V	U	G	A	G	P	G	D	A	G	L	L	T	L	K	G	L	Q	I	Q	Q	A	D	U	U	Y	D						
COBA(9-48)	L	P	A	L	E	K	G	S	W	L	U	G	A	G	P	G	D	P	G	L	L	H	A	A	N	A	N	A	R	Q	D	U	U	Y	D						
COBI(1-40)	L	U	S	G	Y	G	U	G	R	L	I	Q	U	G	T	P	D	P	E	L	L	T	V	K	A	V	K	A	L	Q	Q	A	D	U	U	Y	D				
COBF(1-40)	M	A	E	A	G	M	R	K	I	L	I	G	I	G	S	N	P	E	H	M	T	U	Q	A	I	N	A	N	A	R	Q	E	F	Y	G	Y	F	P	U	L	D
COBJ(1-40)	M	T	G	T	L	Y	U	G	T	Y	U	S	P	W	Q	T	H	E	A	V	A	A	Q	E	A	V	A	A	Q	E	F	Y	G	Y	F	P	U	L	D		
COBL(6-45)	N	S	E	P	A	U	S	P	H	Q	M	T	U	I	G	E	D	G	V	A	G	L	T	E	A	K	R	L	-	A	E	A	P	U	C	Y	A	G			
COBM(1-34)	N	T	U	H	F	I	G	A	G	L	T	E	A	K	R	D	O	L	I	G	R	O	L	T	E	A	K	R	L	-	A	E	A	P	U	C	Y	A	G		

**REGION 2**

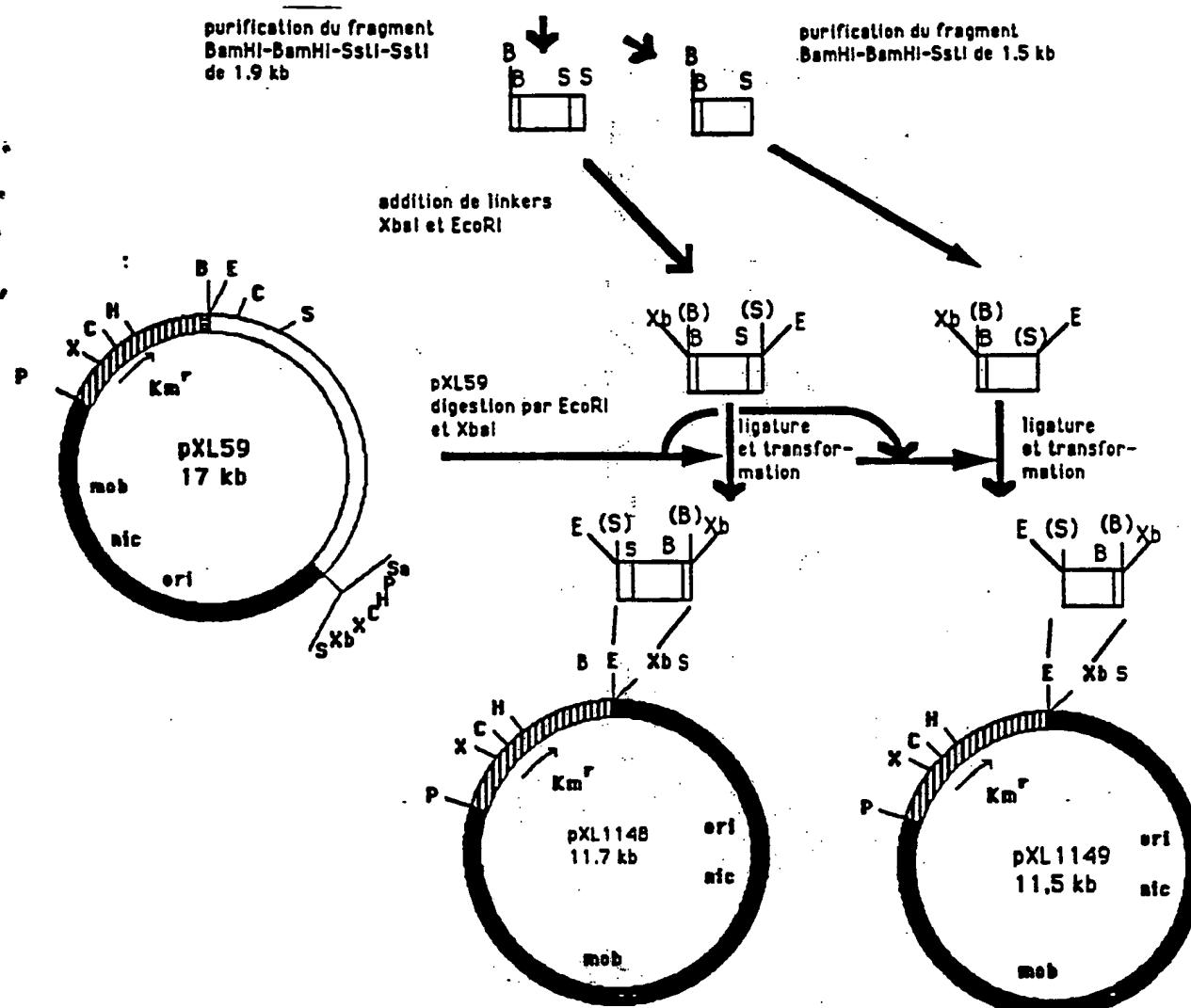
CYSG(288-310)	E	A	Q	K	G	K	R	U	U	R	L	K	G	D	P	F	I	F	G	R	G	C		
COBA(88-110)	L	A	R	A	G	N	R	U	L	K	G	G	D	P	F	F	F	G	R	G	C			
COB(98-120)	H	L	A	R	A	G	R	T	U	A	U	S	E	G	D	P	L	F	Y	G	S	Y	H	
COBF(60-83)	U	T	R	K	D	S	R	T	U	E	F	A	U	P	U	R	R	E	G	U	S	Y	H	
COBJ(68-80)	R	A	A	R	G	U	K	U	C	H	V	S	G	D	P	G	U	F	A	H	A	A	G	
COBL(66-98)	V	A	R	R	G	S	P	V	U	L	A	S	C	D	P	F	F	G	U	G	U	G	U	
COBH(69-91)	A	E	A	E	G	L	D	U	A	R	L	H	S	G	D	L	S	U	W	S	A	U	A	G

**REGION 3**

CYSG(325-342)	U	U	P	G	I	T	A	S	G	S	A	Y	S	G	I	P			
COBA(125-142)	I	U	P	G	I	T	A	G	G	G	A	Y	A	G	I	P			
COB(134-151)	U	-	P	G	-	T	A	H	S	C	C	S	L	A	G	I	P		
COBF(85-102)	S	U	D	D	H	H	A	Q	U	A	G	I	E	A	L	S			
COBJ(110-127)	U	T	P	G	U	T	A	H	L	A	V	A	A	R	I	G	A	P	
COBL(154-171)	R	V	L	T	L	T	S	D	G	A	G	P	R	D	L	A	E	L	
COBM(102-119)	T	W	Y	T	H	T	P	G	U	P	S	F	A	A	S	A			

FIGURE 23

73 / 151



1 ■■■ 2 ■■■■ 3 ■■ 4 ■■

FIGURE 24

74 / 151

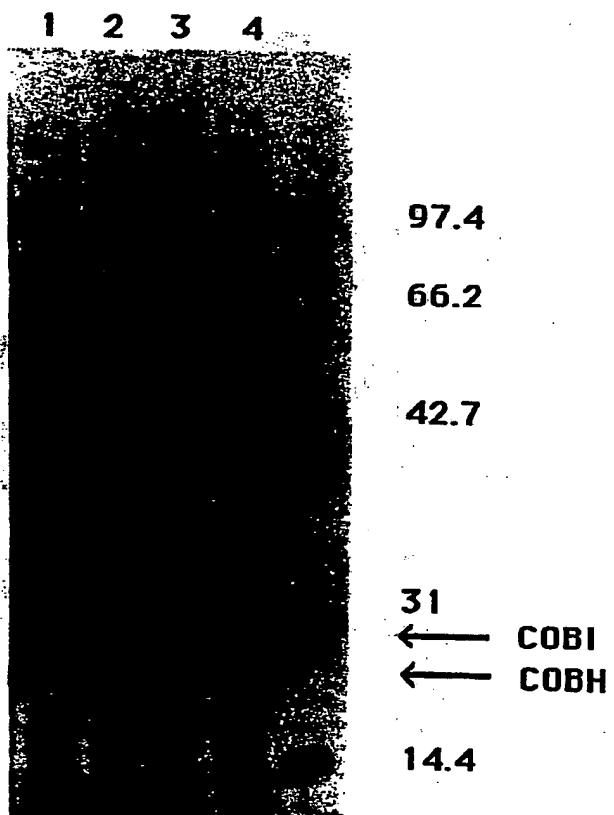


FIGURE 25

FEUILLE DE REMplacement

75 / 151

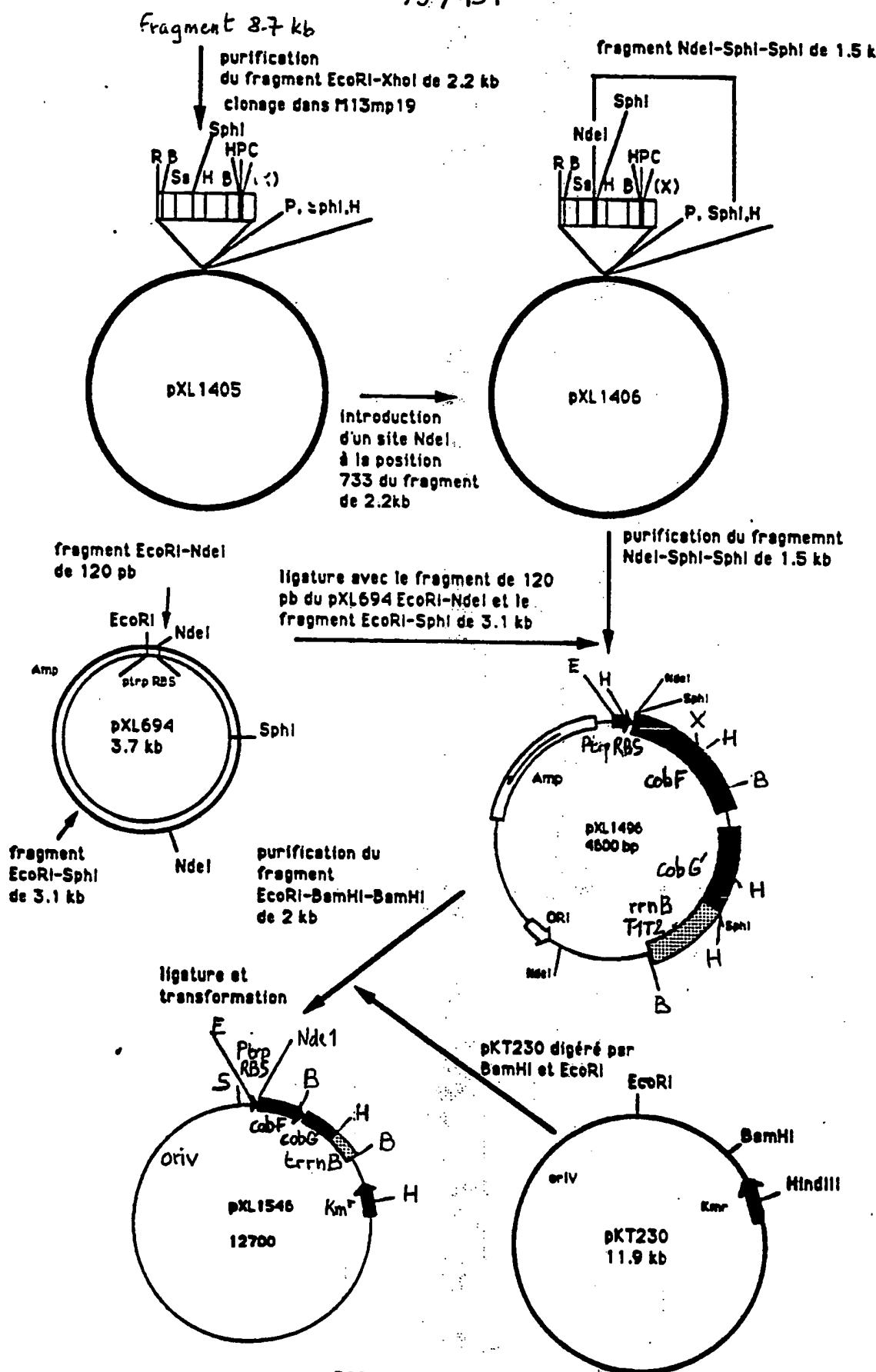


FIGURE 26

PL. 76 / 151

1 2 3

97.4

66.2

42.7

COBF →  
31

14.4

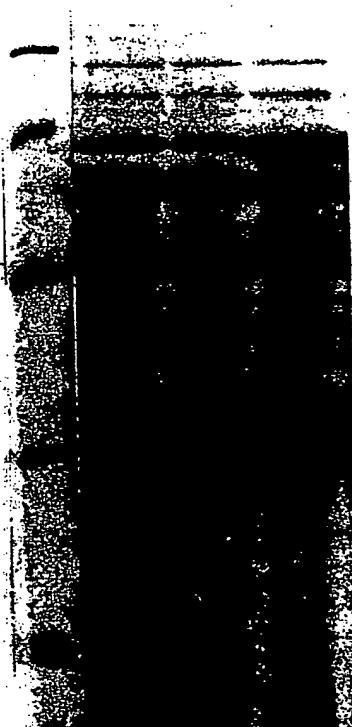


FIGURE 27

PL. 77 | 15 |

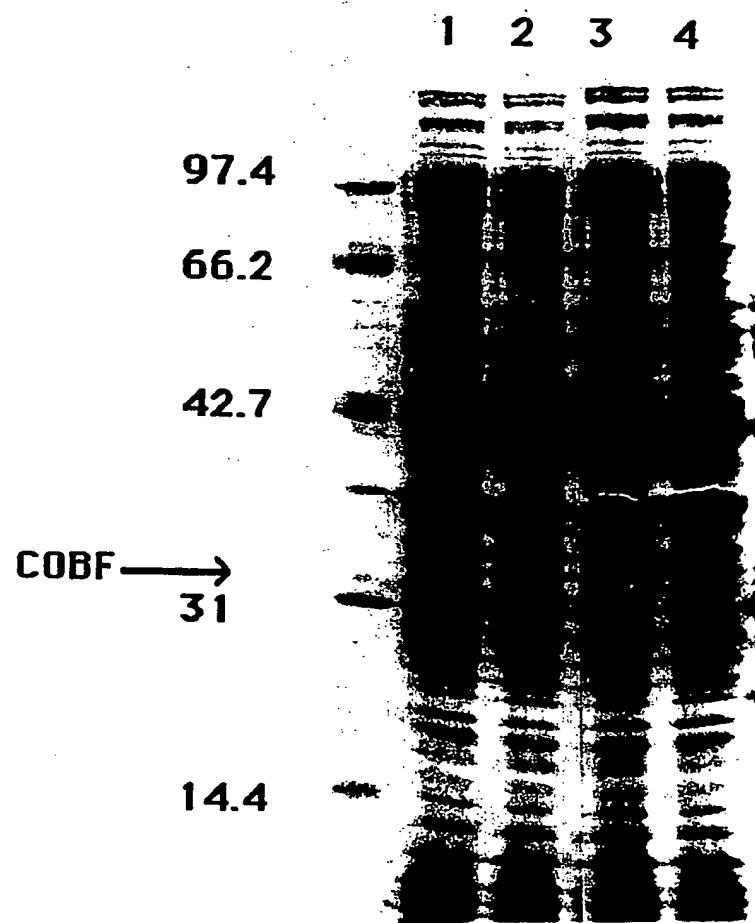


FIGURE 28

FEUILLE DE REMplacement

78 / 151

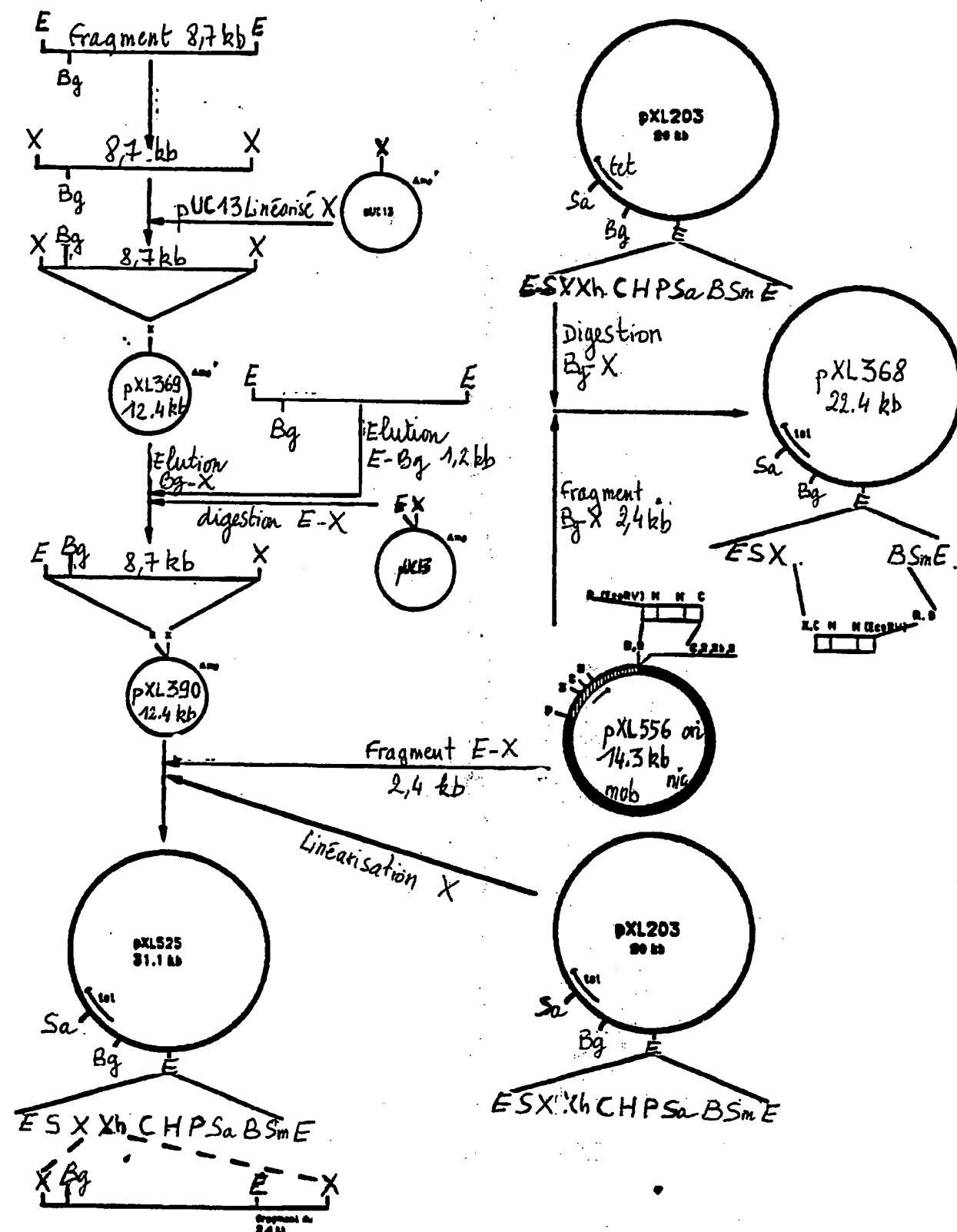


FIGURE 29

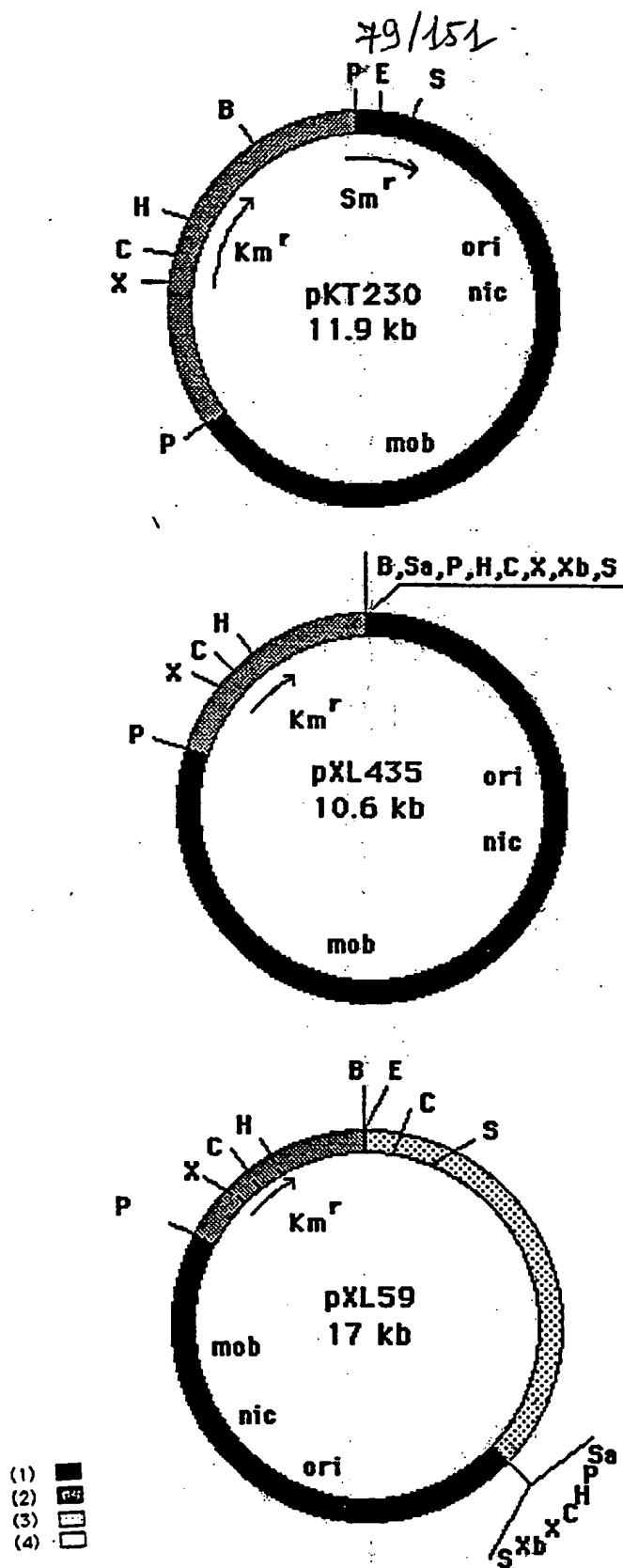


FIGURE 30

80/351

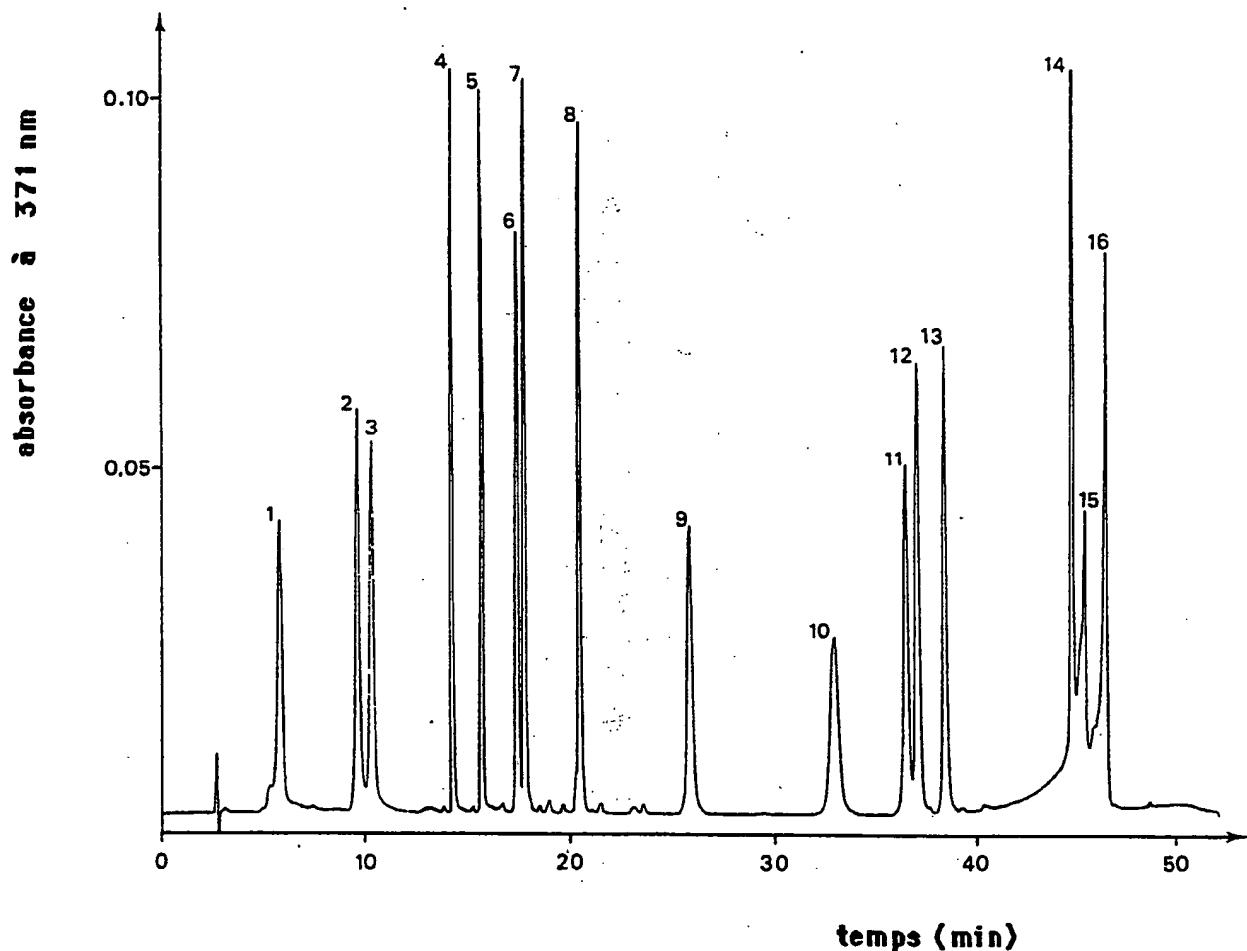


FIGURE 31

PL. 81/15)

10	20	30	40	50	60
GTCGACGGAGT ATGGTCAGGT TCAGGGCTG GTGACCGCTGG AGGACATTCT GGAGGAGATC CAGCTGCTCA TACCAAGTCCA AGTCCCAGAC CACTGGGACC TCCGTAAAGA CCTCCTCTAG					
70	80	90	100	110	120
GTCGGCGATA TCGCCGATGA GCACGACCTC GACATTCAAGG GCGTGCACCA GGAAGCCGAT CAGCCGCTAT AGCGGCTACT CGTGCTGGAG CTGTAAGTCC CGCACCGCGGT CCTTCGGCTA					
130	140	150	160	170	180
GGCTCGATCG TCGTCGATGG CTCGGTGCCTG ATCCCGCATC TCAACCGCGC GCTCGACTGG CCGAGCTAGC AGCAGCTAAC GAGCCACCGC TAGGGCGTAG AGTTGGCGCG CGAGCTGACC					
190	200	210	220	230	240
TCGCTGCCGG ACAGAGGAGGC GACGACGGTG GCCGGCTGG TCATCCACGA GTCCAAGAGC AGCGACGGCC TGCTCCTCCG CTGCTGCCAC CGGCCAGACC AGTAGGTGCT CAGGTTCTCG					
250	260	270	280	290	300
ATTCCGGAGG AGCGCCAGGC TTTCACCTTC CACGGCAAAC GCTTCATCGT GATGAAGCGG TAAGGCTCTCC TGCGGGTCCG GAAGTGGAAAG GTGCCGTTTG CGAAGTAGCA CTACTTCGCC					
310	320	330	340	350	360
GTGAAGAACCG CATTACCAA GCTGCCATC CGTCCGGCGG AAAGGGGTGC TCCGCCGGCG CACTCTTGG CGTAATGGTT CGACCGTAG GCAGGGCGCC TTCTCCCACG AGGCGGCGC					
370	380	390	400	410	420
TGATGGCCGC GATTGCCTCT ACCAGCGGGT CGGCTCGCCG GGGGCTGCCG GCTCGACGGC ACTACCGCGC CTAACGGAGA TGGTCGCCA GCGAGCGGC CCCCGACGGC CGAGCTGCCG					
430	440	450	460	470	480
GAGCGCATGC AGGCCGGCGT CGAGTTCTGG CTTCAGGAGA TCATTGATGG CGCGGTGGCG CTCGCGTACG TCCGGCCGCA GCTCAAGACCA GAAGTCTCT AGTAACIACC CGGCCACCGC					
490	500	510	520	530	540
GGCGACACGG CTCATGCCGG CAAAGCGCT AGAAACGATA CGCACCCGCA TGTGGGACTC CCGCTGTGCC GAGTACGGCC GTTTCCGGGA TCTTGTCTAT CGTGGGGCGT ACACCCCTGAG					
550	560	570	580	590	600
GCCGGTACCA TCGAAGCCCG GCTGATGGCC GGTATGCTGA TGGCTCTCGT TGATGACCTC CGGCCATGGT AGCTTCGGGC CGACTACCGG CCATACGACT ACCGAGAGCA ACTACTGGAG					
610	620	630	640	650	660
GAGCCGTTCA GGGTGGAGG CCTCGATCAG CTTCTTTCG ATGGTCTCGG TGAGCGACAT CTCGCGAAGT CCCACCTTCC GGAGCTAGTC GAAGAAAAGC TACCAAGAGCC ACTCGCTGTA					
670	680	690	700	710	720
TCTTCGTCC CATTTCGCTG TTTGCTTGGC GCCCCCTCGC AGTTAAGAAC CCGGTAATCG AGAAGGCAGG GTAAAACGAC AAACGAACCG CGGGGGAGCG TCAATTCTTG GGCCATTAGC					
730	740	750	760	770	780
CTGGCACGGC GGCGAAAAAT GCCCCCACAA AGCCAGCAAC ATTCCGCTTT GTCAATTCTT GACCGTGCCT CGCGCTTTA CGGGCGTGT TCGGTCGTG TAAGGCAGAA CAGTTAAGAA					
790	800	810	820	830	840
GTTGTGACTC CCCGCAACCC CCATAATGAG CGCCATGAGA CTGATTCAA AATACTTCGA CAACACTGAG GGCGGTTGG GGTATTACTC GCGGTACTCT GAGCTAAGTT TTATGAAGCT					
850	860	870	880	890	900
TCGCATTGCA ACCCGCGCA AGGTGAGGC GCAGCGAGGC CGGGGGCTCC CGTCTGTCAG AGCGTAAGCT TGGGCGCGT TCCAGCTGG CGTCGTCTCG GCGGCCGAGG GCAGACAGTC					
910	920	930	940	950	960
TGGGACGGCT GCGATCAGAA GGGTGTGCAAC CGGGCCCGCG TCGGTGCAA CGCCGAGGGG ACCCCTGCCGA CGCTAGTCIT CCCACACGTG GCGGCCGGGC AGCCAGCGTT CGGGCTCCCC					

FIGURE 32

PL. 82 / 151

970            980            990            1000            1010            1020  
 CAGTACTTCA TGTTCGCTT CGAGCACGTG AAGGAATACA ACAAGGGCTA CAACTCTTC  
 GTCATGAAGT ACAAGACGAA GCTCGTCAC TTCCATTGTT GTTCCCCGAT GTTGAAGAAG  
 1030            1040            1050            1060            1070            1080  
 TCCGGCCTCT CCGACAGCGA GGTCGGCCGC TACCAAGAAGG AAGCGATCAG CGGTCACTCG  
 AGGCCGGAGA GGCTGTCGCT CCAGCGGGCG ATGGCTTCCC TTGCGTAGTG GCCAGTAGCC  
 1090            1100            1110            1120            1130            1140  
 CCCACCTGGA CCGTCGGCGT CAACAAGAAC GCCAAGAACG GCCCGACCCA GTCGCAGACG  
 GGGTGGACCT GGCAGCCGCA GTTGTCTTG CGGTTCTGC CGGGCTGGGT CAGCGTCTGC  
 1150            1160            1170            1180            1190            1200  
 CGGTCTGGCT CTGCGGGCGC CCAGGCCGC ATGCGCGATC CCTTCGGCTT TGTCAGCGAG  
 GCCAGACCGA GACGGCCGCG GGTCCGGCG TAGCGCGTAG GGAAGCCGAA ACAGTCGCTC  
 1210            1220            1230            1240            1250            1260  
 GCGCGGGCGC GATCCGGTCC TCCCGAGCCA CGCCAGCGCA AGCTGAAGAC GCTCGAGGG  
 CGCGCCCGCG CTAGGCCAGC AGGGCTCGGT GCGGTGCGT TCGACTTCTG CGAGCTCCGC  
 1270            1280            1290            1300            1310            1320  
 AAGGCCCTCG AAACGCTTGG TCTCGGAGCC TCGCGACCA CTGCGACAT CAAGGCGGCC  
 TTCCGGAAAGC TTTGCGAACC AGAGCCTCGG AGCCGCTGGT GACGGCTGTA GTTCCGCCGG  
 1330            1340            1350            1360            1370            1380  
 TACAAGGACC TCGTCAAGAA GCATCACCCC GATGCCATG CGGGAGATAG AGGATCGGAA  
 ATGTTCTGG AGCAGTTCTT CGTAGTGGGG CTACCGTTAG CGCCTCTATC TCCTAGCCTT  
 1390            1400            1410            1420            1430            1440  
 GAGCGTTTC GCGCGGTTAT TCAGGCCPAC CAATTGTAA AACAGGCTGG TTTCTGCTAA  
 CTCGAAAAG CGCGCCAATA AGTCCGATG GTTACAATT TTGTCGACC AAAGACGATT  
 1450            1460            1470            1480            1490            1500  
 CAACCCGGAT TAATACAGAA GCACTTTTGC AGCGGAATGC CGGGGTGCCG TCCGGTGGCC  
 GTGGGGCCTA ATTATGTCTT CGTAAAACG TCCGCTTACG CGCCCACGGC AGGCCACCGG  
 1510            1520            1530            1540            1550            1560  
 GCTCTGGAGA CATGATGAGC AAGATTGACC TCGACATTPC CAACCTCCCC GACACCACGA  
 CGAGACCTCT GTACTACTCG TTCTAACTGG AGCTGTAAAG GTTGGAGGGG CTGTGGTGCT  
 1570            1580            1590            1600            1610            1620  
 TTCCGTCCG GGAGGTTTC GGTATTGATA CGGATTTGCG CGTTCCTGCC TATTGGAAGG  
 AAAGGCAGGC CCTCCAAAAG CCATAACTAT GCCTAAACGC GCAAGGACGG ATAAGCTTCC  
 1630            1640            1650            1660            1670            1680  
 GCGACGCCCTA TGTCCCCGGAT CTGGATCGGG ACTACCTCTT CGACCGCGAA ACGACGCTCG  
 CGCTGCCGAT ACAGGGCCTA GACCTAGGCC TGATGGAGAA GCTGGCGCTT TGCTGCGAGC  
 1690            1700            1710            1720            1730            1740  
 CCATTCTCGC AGGCTTCGCC CACAACCGAC GCGTGATGGT GTCGGGCTAT CACGGCACCG  
 GGTAAGAGCG TCCGAAGCGG GTGTTGGCTG CGCACTACCA CAGCCCGATA GTGCCGTGGC  
 1750            1760            1770            1780            1790            1800  
 GCAAGTCCAC CCATATCGAG CAGGTGCCG CGCGCCTCAA CTGGCCGTGC GTGCCGTCA  
 CGTTCAGGTG GGTATAGCTC GTCCAGCGGC CGCGGGAGTT GACCGGCACG CACGCGCAGT  
 1810            1820            1830            1840            1850            1860  
 ACCTCGATAG CCATGTCAAG CGTATCGAC TCGTCGGCAA GGACCGCGATC GTCGTCAAGG  
 TGGAGCTATC GGTACAGTCG GCATAGCTGG AGCAGCCGTT CCTGCGCTAG CAGCAGTTCC  
 1870            1880            1890            1900            1910            1920  
 ACGGCCTGCA GGTCAACCGAA TTCAAGGAGC GCATCGTGC CTGGGCCTAC CAGCACAATG  
 TGCCGGACGT CCAGTGGCTT AAGTCTCTGC CGTAGGACGG GACCCGGATG GTCGTGTAC

FIGURE 32.1

FEUILLE DE REMPLACEMENT

PL. 83 / 15 /

1930	1940	1950	1960	1970	1980
TCGGCCTCGT	CTTCGACGAA	TACGATGCCG	GCGCCCGGGA	CGTCATGTTC	GTCATCCAGC
AGCGCGAGCA	GAAGCTGCTT	ATGCTAAGGC	CGGGGGCCCT	GCAGTACAAG	CAGTAGGTGCG
1990	2000	2010	2020	2030	2040
GCGTGTGGA	ATCCCTCCGGC	CGCCCTGACGC	TGCTGACCA	GAGCCGTGTC	ATCCGTCCGC
CGCACGACCT	TAGGAGGCGC	CGGGACTGCG	ACGAGCTGGT	CTCGGCACAG	TAGGCAGGCG
2050	2060	2070	2080	2090	2100
ACCCGGCCCTT	CCGCTGTTT	GCGACCGCCA	ACACCGTCGG	CCTCGGCAC	ACGACCGGCC
TGGGCGGAA	GGCGGACAAA	CGCTGGCGGT	TGTGGCAGCC	GGAGCCGCTG	TGCTGGCCGG
2110	2120	2130	2140	2150	2160
TCTATCACGG	CACCGCAGCAG	ATCAACCCAGG	CGCAGATGGA	CCGCTGGTTC	ATCGTCACCA
AGATACTGCC	GTGCGTCGTC	TAGTTGGTCC	CGCTCTACCT	GGCGACCCAGC	TAGCAGTGGT
2170	2180	2190	2200	2210	2220
CGCTGAACTA	CCTGCCGCAC	GACAAGGGAG	TCGACATCGT	CGCCGCCAAG	GTCAAGGGCT
GGCAGTTGAT	GGACGGCGTG	CTGTTCTTC	AGCTGTAGCA	GGGGCGGTTC	CAGTTCCCCGA
2230	2240	2250	2260	2270	2280
TCACCGCCGA	CAAGGGCCGC	GAGACCGTCT	CCAAGATGGT	ACGTGTGCGC	GACCTCACGC
AGTGGCGGCT	TTTCCCAGCG	CTCTGGCAGA	GGITCTACCA	TGACACAGCGG	CTGGAGTGGC
2290	2300	2310	2320	2330	2340
GCGCAGCCTT	CATCAATGGC	GATCTCTCGA	CTGTCATGAG	CCCGCGTACG	GTCATCACCT
CGCGTCGGAA	GTAGTTACCG	CTAGAGAGCT	GACAGTACTC	GGGCGCATGC	CAGTAGTGGA
2350	2360	2370	2380	2390	2400
GGGCCGAGAA	CGCCACATC	TTCGGCGACA	TCGCTTCGCG	CTTCCGCGTG	ACCTTCCTCA
CCC GGCTCTT	GC GGGGTGTAG	AAGCCGCTGT	AGCGAAAGCG	GAAGGCGCAC	TGGAAGGAGT
2410	2420	2430	2440	2450	2460
ACAAGTGCAGA	CGAGCTGGAG	CGGGCGCTGG	TCGCGAGCA	CTACCAGCGC	GCCTTCGGCA
TGTTACCGCT	GCTCGACCTC	GCCCCGGGACC	AGCGGCTCGT	GATGGTGC	CGGAAGCCGT
2470	2480	2490	2500	2510	2520
TCGAGCTGAA	GGAAATGCGCT	GCCAACATCG	TGCTCGAACG	CACCGCCTGA	TCCCACGGCC
AGCTCGACTT	CCTTACCGCA	CGGTTGTAGC	ACGAGCTTCG	GTGGCGGACT	AGGGTGCCGG
2530	2540	2550	2560	2570	2580
TGCCGTCCCC	TTTGGGAGGG	CGGGTCATGA	CCCTGTGGCA	AACCGGATGA	CGCCCCACTG
ACGGCAGGGG	AAACCCCTCCC	GCCCCAGTA	CGCACCCGT	TTGGCCTACT	GGGGGGTGAC
2590	2600	2610	2620	2630	2640
GGGCGCCGTC	GCCTCTGGCT	GAAGAAGGAA	CTGTCGTGAG	CTCGAAATTG	AAGGCAAAGC
CCC CGGGCAG	CGGAGACCGA	CTTCTTCCTT	GACAGCACTC	GAGCTTAAGC	TTCCGTTTCG
2650	2660	2670	2680	2690	2700
CAACCAACCGC	CGAGAAATGCT	GCGGAACCGT	TCAAGCGGGC	GCTTCGGC	TGCATCCGAT
GTGCGCGC	GCTCTTACGA	CGCTTGGCA	AGTCGCCCCG	CGAAAGGCCG	ACGTAGGCTA
2710	2720	2730	2740	2750	2760
CGATCGCGGG	CGATGCCGAG	GTGGAAAGTGG	CCTTCGCCAA	CGAGCGGCCG	GGCATGACCG
GCTAGCGCCC	GCTACGGCTC	CACCTTCAGC	GGAAAGCGGT	GCTCGCCGGC	CCGTACTGGC
2770	2780	2790	2800	2810	2820
GCGAACGCAT	CCGTCTGCCG	GAACCTTCGA	AGCGCCCGAC	CCTGCAGGAA	CTTGGCGTGA
CGCTTGCCTA	GGCAGACGGC	CTTGAAAGGT	TGCGGGCTG	GGACGTCCTT	GAACGGCACT
2830	2840	2850	2860	2870	2880
CCCGCGGGCT	CGGTGACAGC	ATGGCGCTGC	GCAAGGCCTG	TACGCATGCG	CGGATCCAGC
GGGCGCCCGA	GCCACTGTGCG	TACCGCGACG	CGTCCGGAC	ATGCGTACGC	GCCTAGGTGCG

FIGURE 32.2

PL. 84 / 151

2890	2900	2910	2920	2930	2940
GCACCATGTC	GCCGCAAGGG	GCGGACGCC	GCGCGATCTT	CGATCGGGTG	GAGCAGGCTC
CGTGGTACAG	CGGCCTTCCC	CGCCTGCCGG	CGCGCTAGAA	GCTACGCCAC	CTCGTCCCAG
2950	2960	2970	2980	2990	3000
GTGTCGAGGC	GATCGGGTCG	TTGCGCATGG	CGGGTGTGCG	CAAGAACCTC	AACGTATGC
CACAGCTCCG	CTAGCCCAGC	AACCGCTAAC	GCCCCACAGCG	GTTCTGGAG	TTGCAAGTACG
3010	3020	3030	3040	3050	3060
TCGAAGAGAA	ATACGCCAAG	GCGAATTTCG	CAACGATCGA	GCGCCAGGCG	GACGCCGCC
AGCTTCTCTT	TATGCGGTTC	CGCTTAAGC	GTTGCTAGCT	CGCGGTCCGC	CTGCGCGGCC
3070	3080	3090	3100	3110	3120
TCGGCGAGGC	CGTAGCGC	GCTGGTGCCTG	AGAACGCTGAC	GGGCCAGAAC	CCGCCGGCGT
AGCCCGCTCCG	GCATCGCGAC	GACCACCGC	TCTTCGACTG	CCCCTGCTTC	GGCGGCCGCA
3130	3140	3150	3160	3170	3180
CTGCCGGCAA	GGTGCTCGAC	CTCTGGCGCG	AGTTCATCGA	GGGCAAGGCT	GCCGGCGACA
GACGGCCGTT	CCACGAGCTG	GAGACCGCGC	TCAAGTAGCT	CCCCTTCCGA	CGGCCGCTGT
3190	3200	3210	3220	3230	3240
TTGAGCACCT	GTCGTCGACG	ATCAACAAAC	AGCAGGCCCTT	TGCCCCGGTC	GTTCGCGACA
AACTCGTGGA	CAGCAGCTGC	TAGITGTTGG	TGTCGGGAA	ACGGGCCCCAG	CAAGCGCTGT
3250	3260	3270	3280	3290	3300
TGCTGACCTC	GATGGAAGTC	GCCGAGAAAT	ACGGTGCAGA	CGACAAACGAG	CCGGACGAGC
ACGACTGGAG	CTACCTTCAG	CGGCTCTTAA	TGCCACTGCT	GCTGTTGCTC	GGCCTGCTCG
3310	3320	3330	3340	3350	3360
AGGAAAGCGA	GACCGACGAA	GACCAGCCGC	GCAGGCCAGGA	GCAGGACGAG	AACGCCAGCG
TCCCTTCGCT	CTGGCTGCTT	CTGGTCGGCG	CGTCGGTCTC	CGTCCTGCTC	TTGCGGTCCG
3370	3380	3390	3400	3410	3420
ACGAGGAAGC	CGGGCAGCGAT	GCCGCACCCG	CCGACCGAGAA	CCAGGCTGCC	GAAGAGCAGA
TGCTCCTTCG	GCCGCTGCTA	CGGCCTGGGC	GGCTGCTCTT	GGTCCGACGG	CTTCTCGTCT
3430	3440	3450	3460	3470	3480
TGGAAAAGG	CGAGATGGAC	GGCCCGGAGA	TCTCCGACGA	CGATCTCCAG	GACGAAGGCC
ACCTTCTTCC	GCTCTACCTG	CGCGCCCTCT	AGAGGTGCT	GCTAGAGGTC	CTGCTTCCG
3490	3500	3510	3520	3530	3540
ACGAGGACAG	CGAAACGCC	GGCGAGGTCA	AGCGTCCGA	CCAGCCCTTC	GGCGACTTCA
TGCTCCTGTC	GCTTTCGGGG	CCGCTCCAGT	TGCGAGGCTT	GGTCCGGAAAG	GGCGCTGAAGT
3550	3560	3570	3580	3590	3600
ACGAGAAGGT	CGACTACGCC	GTCCTCACCC	GCGAGTTGCA	CGAGACGATT	GCCTCGGAAG
TGCTCTTCCA	GCTGATGCCG	CAGAAGTGGG	CGCTCAAGCT	GCTCTGCTAA	CGGAGCCCTC
3610	3620	3630	3640	3650	3660
AGCTTTCGCA	CGAGGCCGAG	CTCGACCGGC	TGCGCGCTT	CCTCGACAAG	CAGCTTGCCC
TCGAAACGCT	GCTCCGGCTC	GAGCTGGCCG	ACGCGCGAA	GGAGCTGTT	GTGAAACGGG
3670	3680	3690	3700	3710	3720
ATCTTCAAGG	CGCGGTCGGC	CGCCTTGCCA	ACCGGTGCA	GCGCCGCCCTG	ATGGCGCAGC
TAGAAGTTCC	GCGCCAGCCG	GGCGAACGGT	TGGCCGACGT	CGCGGCGGAC	TACCGCGTCG
3730	3740	3750	3760	3770	3780
AGAACCGCTC	CTGGGAGTTC	GATCTCGAAG	AGGGGTATCT	CGATTCGGCG	GGGCTTCAGC
TCTTGGCGAG	GACCCCTCAAG	CTAGAGCTTC	TCCCCATAGA	GCTAAGCCGC	GCCGAAGTCG
3790	3800	3810	3820	3830	3840
GCATCATCAT	CGATCCGATG	CAGCCGCTT	CCTTCAGCG	CGAAAAGGAC	ACCAACTTCC
CGTAGTAGTA	GCTAGGCTAC	GTCGGCGAAA	GGAAAGTCGC	GCTTTCTG	TGGTTGAAGG

FIGURE 32.3

PL. 85 / 111

3850	3860	3870	3880	3890	3900
GGCATACCGT	CGTGACGCTG	CTGATCGACA	ATTCCGGCTC	GATGCGCGGC	CGTCCGATCA
CGCTATGGCA	GCACTGCGAC	GACTAGCTGT	TAAGGCCAG	CTACGCGCCG	GCAGGCTAGT
3910	3920	3930	3940	3950	3960
CGGTGCGCGC	CACCTGCGCC	GATATCCTGG	CGCGCACGCT	CGAGCGCTGC	GGCGTCAGG
GCCAACGGCG	GTGGACGCGG	CTATAGGACC	GCGCGTGCAG	GCTCGCGACG	CCGCAGTTCC
3970	3980	3990	4000	4010	4020
TGGAGATCCT	CGGTTTAC	ACCAAGGCGT	GGAAAGGGTGG	GCAGTCACGC	GAGAAGTGGC
AGCTCTAGGA	GCCAAAATGG	TGGTCCGCA	CCTTCCCAC	CGTCAGTGC	CTCTTCACCG
4030	4040	4050	4060	4070	4080
TGGCCGGCGG	CAAGCC	CAG	GCCCCGGGTC	GCCTCAACGA	CCTGCGACAC
ACCGGCCGCGC	TTTCGGTGT	CGGGGCCAG	CGGAGTTGCT	GGACGCTGTG	TAGCAGATGT
4090	4100	4110	4120	4130	4140
AGCTGCCGA	CGCTCCGTGG	CGCCGGGCAC	GACGCAATCT	CGGCCTGATG	ATGCGGGAAAG
TCAGACGGCT	GCGAGGCACC	CGGGCCCCGTG	CTGCGTTAGA	GCCGGACTAC	TACGCCCTTC
4150	4160	4170	4180	4190	4200
GCCTGCTCAA	GGAAAACATC	GACGGCGAGG	CGTTGATTTG	GGCGCATGAG	CGGCTGATGG
CGGACGAGTT	CCTTTTGTAG	CTGCCGCTCC	GCAACTAAAC	CGCGTACTC	GCCGACTACC
4210	4220	4230	4240	4250	4260
CGCGCGCGA	ACAGCGGCAC	ATCCTGATGA	TGATTTCGGA	CGGCGCGCCG	GTCGACGACT
CGGCCGCGCT	TGTCGCCGCG	TAGGACTA	ACTAAAGCCT	GCCGCGCGGC	CAGCTGCTGA
4270	4280	4290	4300	4310	4320
CGACGCTGTC	GGTCAATCCA	GGAAAATATC	TGGAGCGTCA	CCTGCGCGCG	GTCATCGAGC
GCTGCGACAG	CCAGTTAGGT	CCTTGATAG	ACCTCGCAGT	GGACGCGCGC	CAGTAGCTCG
4330	4340	4350	4360	4370	4380
AGATCGAAAC	GCGCTCGCCG	GTGGAACATC	TGGCGATCGG	TATCGGCCAC	GACGTGACGC
TCTAGCTTTG	CGCGAGCGGC	CACCTTGACG	ACCGCTAGCC	ATAGCGGGTG	CTGCACTGCG
4390	4400	4410	4420	4430	4440
GCTACTATCG	CCGTGCCGTC	ACCATCGTC	ATGCCGATGA	GCTTGCCGGC	GCGATGACCG
CGATGATAGC	GGCACGGCAG	TGGTAGCAGC	TACGGCTACT	CGAACGGCCG	CGCTACTGGC
4450	4460	4470	4480	4490	4500
AACAGCTGGC	CGCACTCTTC	GAGGACGAAA	GCCAGCGCCG	CGGTTCTCG	CGTCCTCGCC
TTGTCGACCG	GCGTGAGAAG	CTCCCTGCTT	CGGTCGCGGC	GCCAAGAAC	GCAGAACCGG
4510	4520	4530	4540	4550	4560
GGGCCGGGTG	ATGCTTCCCC	CTTGGGGCG	GTGGAACATC	GCCTCCGAGC	TGCCAATCGG
CGCGGCCAC	TACGAAGGGG	GAACCCCCGC	CACCTTGATG	CGGAGGCTCG	ACGGTTAGCC
4570	4580	4590	4600	4610	4620
CACCTGCAAG	CATCGCTGGC	GGCCGAAGTC	AATTCACGGA	CATAGTTTTA	CAGTCTACCA
GTGGACGTGC	GTAGCGACCG	CCGGCTTCAG	TTAATGCTT	GTATCAAAT	GTAGATGGT
4630	4640	4650	4660	4670	4680
AGCTACCATG	CGTGGCGGGC	TCACTTTGAG	CGCACGCCG	GTCAATTCCCG	ATGCCCCCTG
TCGATGGTAC	GCACCGCCCG	AGTGAACATC	GCGTGCAGGC	CAGTAAGGGC	TACGGGGGAC
4690	4700	4710	4720	4730	4740
AAGGTACTTC	TCTTGATGCT	TGGCCGGGT	CTCCCTAGCCC	TTTCCTCCCT	GGCTTCGGCC
TTCCATGAAG	AGAACTACGA	ACCGGCAGCA	GAGGATCGGG	AAAAGGGAGGA	CCGAAGCCGG
4750	4760	4770	4780	4790	4800
TGCCCGGC					
ACGGGCCG					

FIGURE 32.4

PL. 86/151

10 20 30 40 50 60  
 GAGCTCATAG AGCAGTTCCCT CGATCGACTT CAGCAGTCGC ATGAAATCCA TGCCGTGCTC  
 CTCGAGTATC TCGTCAAGGA GCTAGCTGAA GTCGTCAGCG TACTTTAGGT ACGGCACGAG  
 70 80 90 100 110 120  
 CCCTTGCTTC TATGCGTGGC ACGACCGCGC GCCGGGGCCG ATGCCGGTCA GTCGCGCAGA  
 GGGAACGAAG ATACGCACCG TGCTGGCGCG CGGGCCCCGGC TACGGCCAGT CAGCGCGCTC  
 130 140 150 160 170 180  
 CGCAGCTCGT CGGTACGCAT CTGCAGCATC TCCAGCGTCG ACAGGAAGCT CATGCCGAGC  
 GCGTCGAGCA GCCATGCTTA GACGTCGTAG AGGTCGCAGC TGTCCCTCGA GTACGGCTCG  
 190 200 210 220 230 240  
 AGGCTCTGAT CGAGCTTGCC CTTGGCTGCG ACCGTTGCGC CGATGTTGCG GCGGGGTGATC  
 TCCGAGACTA GCTCGAACGG GAACCGACGC TGGCAACGGC GCTACAACGC CGCCCACTAG  
 250 260 270 280 290 300  
 GGGCCGATCG AGATCTCCTG AAGCATCACG GGGGCTGCTC GGGCCCCGGC ATTGGCTGTC  
 CCCGGCTAGC TCTAGAGGAC TTCGTAGTGC CCCCCGACGGA CCCGGGGCCGG TAACCGACAG  
 310 320 330 340 350 360  
 ATGACCGTGA CGATAAAGIT GAGGTGCGCC GGGTCGAGGC CGATCTTTC CGCATCTTCA  
 TACTGGCACT GCTATTCTCAA CTCCAACCGG CCCAGCTCCG GCTAGAAAAG GCGTAGAAAGT  
 370 380 390 400 410 420  
 TAGGTGAGCG CGATGTTGCT GGCGCCGCTA TCGACCAGCA TGGTGTGTC CTTGCGCTCG  
 ATCCACTCGC GCTACAAACGA CGCGGGCCAT AGCTGGTEGT ACCGACTACAG GAACGGCAGC  
 430 440 450 460 470 480  
 ACCGTGCGAG TGGTCTCGAA ATGACCGTTC AGCATTCTC GCAGCACCAC TTCCCTGCTG  
 TGGCAGCGTC ACCAGAGCTT TACTGGCAAG TCGTAGAAGA CGTCTGGTG AAGGACGACA  
 490 500 510 520 530 540  
 CCCTCGCTGT CAGTGTGAT GGTGGCGCGG CCGGGGATGA GGGCCGGCGAG CAGGCGGTTA  
 GGGAGCGACA GTCACTACTA CCACCGCGCC GGGCCCTACT CCGGGCCGCTC GTCCGCCAAT  
 550 560 570 580 590 600  
 CCGAAGCCCT CCAACTCGAA GCGGTAGACA TAGGCCGAGA CCAGCGCCAG AACGACGAAG  
 GGCTTCGGGA GGTTGAGCTT CGCCATCTGT ATCCGGCTCT GGTCGGCGTC TTGCTGCTTC  
 610 620 630 640 650 660  
 AGCCAGATGG CGATCTGACG CAGGCCCTCG CCGAAGCGGT GGGCGCTCTG CAGGATGCCG  
 TCGGTCTACC GCTAGACTGC GTCCGGAAAGC GGCTTCGCCA CGGCCGAGAC GTCCCTACGGC  
 670 680 690 700 710 720  
 GCGCCGATCA GCGTGGCGAT GGCGCCGAGC GAGACCAAGT GCGCGAACTG GTCATTGGCA  
 CGCGGCTAGT CGCACCGCTA CGCGGGCTCG CTCTGGTCAA CGGGCTTGAC CAGTAACCGT  
 730 740 750 760 770 780  
 AGCCCCATGG TGGGGCCGGT GTCTGGTTG ATGATCAGCA GGATGAGGCC GATGGCCAGG  
 TCGGGGTACC ACGCCGGCCA CAGCACCAAC TACTAGTCGT CCTACTCCGG CTACCGGTCC  
 790 800 810 820 830 840  
 ATCGAGAGCA GGATGGCAAG ACGGGTCTATG CTTCGGCGCG TTCCCTCGCC ATGCGCTG  
 TAGCTCTCGT CCTACCGTTC TCCCCAGTAC GAAGCGGCC AAGGGAGCGG TACGCGCACG  
 850 860 870 880 890 900  
 GTCGGGTTTC GCGCCCGCGC TTGCGTTCGA CGGTCTCAAG CGGTGCAAGG AACGCGCTCA  
 CAGCCCAAAG CGCGGGCGCC AACGCAAGCT GCGAGAGTTC GGCACGTCCG TTGCGCGAGT  
 910 920 930 940 950 960  
 TGATCGCGCG GCGGTGGCA TCGGTATAAGA GCGTCCAGCG TCCGACTTCG TCGCGGGTAC  
 ACTAGCGCGC CGCAAGCCGT AGCCATCTT CGCAGGTGCG AGGCTGAAGC AGCGCCCATG

FIGURE 33

PL. 87/151

970	980	990	1000	1010	1020
GGCCGCAGCC	GAAACAGTAG	CCGGTCTGT	CATCGATCGA	ACAGACGAGA	ATGCAGGGAG
CCGGCGTCGG	CTTGTCTATC	GGCCAGAAC	GTAGCTAGCT	TGCTGCTCT	TACGTCCCTC
1030	1040	1050	1060	1070	1080
ATTCCATGGG	CGTGCTCATG	TTTCCCTTGA	TATATCGATG	TTTCAAACCG	TCAGCGCAAG
TAAGGTACCC	GCACCGAGTC	AAAGGGAACT	ATATAGCTAC	AAAGTTGGC	AGTCGGCTTC
1090	1100	1110	1120	1130	1140
GGCACCGAGC	ACGGCGATTT	CGGTCAGTIG	CTGCCGCGCC	CCGATCGTGT	CGCCCGTTG
CCGTGGCTCG	TGCCGCTAA	GCCAGTCAAC	GACCGAGCGG	GGCTAGCACA	GGGGCAAAC
1150	1160	1170	1180	1190	1200
TCCGCCGATC	TTGCGCATCG	CGAGCCGAGC	GAAGGCCCTTG	ACCGTGGCAA	GGAATGCGAC
AGGGGGCTAG	AACGGTAGC	GGTCGGCTCG	CTTCGGGAAC	TGGCACCGTT	CCTTACGCTG
1210	1220	1230	1240	1250	1260
GAGGCCGCG	ATGACGCCGA	GCGCCGGAC	CTGCCGAGA	TAGAAGAGCA	GCATTGCGAC
CTCGCGCGC	TACTGCCGCT	CGCGGCCCTG	GACGCCGCTCT	ATCTTCTCGT	CGTAACGCTG
1270	1280	1290	1300	1310	1320
AAGAAGTCCG	AAGGCAAGCG	CGAAGCCGCT	GGCCGCCGGT	TCCGGCTCGC	CAGCCGAGGC
TTCTCAGGC	TTCCGTTCGC	GCTTCGCGCA	CCGGCGGGCA	AGGCCGAGCG	GTCGGCTCCG
1330	1340	1350	1360	1370	1380
CGCGACGCCG	CTGCTGCGCG	CCGGCGGAAG	CGACGACCAG	TGCCAGACCA	TGGCGGGCGCG
GGCGCTCGC	GACGACGCGC	GGCCGCCCTC	GCTGCTGGTC	ACGGTCTGGT	ACCGCCGCGC
1390	1400	1410	1420	1430	1440
GCTGAGGCAC	GCTGCGCCAA	GGATCGCAT	GGCGGCCCCC	AGCGGGCAGA	AGAGCGGCAG
CGACTCCGTG	CGACGGGTT	CCTAGCGGTA	CCGCCCGGGG	TGCCCGCTTT	TCTCGCCGTC
1450	1460	1470	1480	1490	1500
GATCGAGGCG	AACGCCGAGA	CGCGCAGGCC	GAAGGAGAGG	ATGAGGGCGA	CGGCCGCATA
CTAGCTCCGC	TTGCGGCTCT	GGCGTCCCGG	CTTCCTCTCC	TACTCCCCGCT	GCCGGCGTAT
1510	1520	1530	1540	1550	1560
GGTGCCGATG	GGGCTGTCT	TCATGATCGC	AAGGCCGCT	TCGCGGTGCG	GACCGCCGCC
CCACGGCTAC	GCCGACAGGA	AGTACTAGCG	TTCGCGCGA	AGCGGCCAGCG	CTGGCGGGCGG
1570	1580	1590	1600	1610	1620
AAAGCCATCG	GCCGTGTCGC	CAAGCCCGTC	TTCGTGCAGT	GGCCCGGTGA	CAAGCGCTG
TTTGGTAGC	GGGACACAGCG	GTTCGGGAG	AAGCACGTC	CGCGGCCACT	TTTCGCGGAC
1630	1640	1650	1660	1670	1680
GATGGCGACG	ACGACAAAGG	CGGCAAAGAG	CGAGCTCACC	TGCAGCGCCA	TGAGGGCCAT
CTACCGCTGC	TGCTGTTTCC	GGCGTTCTC	GCTCGAGTGG	ACGTGCGGGT	ACTCCCGGTA
1690	1700	1710	1720	1730	1740
GGCGACGGCC	GCCGATGGCA	GTGCGATCGC	CAGGGCGCG	AACGGGAAGG	CGCGCACCGC
CCGCTGCCGG	GGGCTACCGT	CACGCTAGCG	GTCCGGCGC	TTGCCCTTCC	GCGCGTGCCG
1750	1760	1770	1780	1790	1800
ACGGCTCAAG	CGCCCCGTAT	AACCTTCGAA	ATGACGCCA	GGCATCGGGA	TGCGGCTGAG
TGCGGAGTTC	GGGGCGAGTA	TTGGAAGGT	TACTGCCGT	CCGTAGCCCT	ACGCCGACTC
1810	1820	1830	1840	1850	1860
AAAGCCGATC	GACCGCGCCA	CATCGTCACA	GAAATGCCA	ACGAAGGCCA	TGGCTCCCTCC
TTTGGCTAG	CTGGCGCGGT	GTAGCAGTGT	CTTTCAGCGT	TGCTTCGGGT	ACCGAGGAGG
1870	1880	1890	1900	1910	1920
AAGGTTGCGG	CCATTGACCC	GGCGCGTGC	AAACTCGCCG	ACTGCGGCAGA	GTCTCGCAAG
TTCCAACGCC	GGTAACCTGGG	CCGGCGACGG	TTTGAGCGGC	TGACGCCGCT	CAGAGCGTTC

FIGURE 33.1

FEUILLE DE REMPLACEMENT

PL. 88 / 151

1930	1940	1950	1960	1970	1980
CCGGGCGGGC	GCACCCGCGA	GGGCCGCGA	CACTTTCCC	AGACCTTCA	TAGGCCGTCT
GGCCCGCCCG	CGTGGCGCT	CCCGCGCGT	GTAAAAGGG	TCTGGAAAGT	ATCCGGCAGA
1990	2000	2010	2020	2030	2040
GCGACCGCTC	GCGGATCGAG	ACGGCGACGC	CGATTGGCGC	AAATGTCGTT	GCCCCAATT
CGCTGGCGAG	CGCCTAGCTC	TGCCGCTGCG	GCTAACCGCG	TTTACAGCAA	CGGGCTAAA
2050	2060	2070	2080	2090	2100
TCGGCGCCCT	CTATGAGGGG	CGTAGATAGA	GCTTCACGAT	GATGCAAGGA	TTCCCTCCCAT
AGCCGCGGGG	GATACTCCCC	GCATCTATCT	CGAAGTGCTA	CTACGTTCT	AAGGAGGGTA
2110	2120	2130	2140	2150	2160
GAGTGCCAGC	GGCCTGCCGT	TTGATGATTT	TCGCGAATTG	TTGCGCAACC	TGCCGGGCC
CTCACGGTCG	CCGGACGGCA	AACTACTAAA	AGCGCTTAAC	AACGCGTTGG	ACGGCCCGGG
2170	2180	2190	2200	2210	2220
GGATGCGGCA	GCCCTCGTTG	CCGCGCGGGA	GCGGGACGCC	CAGCTGACGA	AGCCGCGGGG
CCTACGCCGT	CGGGAGCAAC	GGCGCGCCCT	CGCCCTGCGG	GTCGACTGCT	TGCCGGGCC
2230	2240	2250	2260	2270	2280
CGCGCTCGGC	CCCTCGAGG	AAATGCCCTT	CTGGCTCGCC	GCCTGGACGG	GCAAGGCGCC
GCGCGAGCCG	CGGGAGCTCC	TTAGCGGAA	GACCGAGCGG	CGGACCTGCC	CGTTCCGCGG
2290	2300	2310	2320	2330	2340
GGTGGTCAAC	CGGCCGCTGG	TGGCGATTT	TGCCGCAAC	CACGGCTCA	CCCGCCAGGG
CCACCACTTG	GCGGGCGACC	ACCGCTAGAA	ACGGCCGTG	GTCGCCAGT	GGCGGGTCCC
2350	2360	2370	2380	2390	2400
GGTGACCCCG	TTCCCGTCAT	CCGTCAACCGC	ACAGATGGTC	GAGAATTTG	CCGCCGGTGG
CCACTGGGGC	AAGGGCAGTA	GGCAGTGGGG	TGTCTACCAAG	CTCTTAAAC	GGCGGCCACC
2410	2420	2430	2440	2450	2460
CGCTGCGATC	AACCAGATCT	GCGTCAGCCA	CGACCTCGGG	CTGAAGGTCT	TCGACCTCGC
GCGACGCTAG	TTGGTCTAGA	CGCAGTCGGT	GCTGGAGCCC	GACTTCCAGA	AGCTGGAGCG
2470	2480	2490	2500	2510	2520
ACTCGAAATAC	CCGACCGGTG	ATATCACCGA	GGAAGCCGCG	CTGTCCGAGC	GCGATTGCGC
TGAGCTTATG	GGCTGGCCAC	TATAGTGGCT	CTTCCGGCGC	GACAGGCTCG	CGCTAACCGG
2530	2540	2550	2560	2570	2580
CGCGGACCATG	GCCTTTGGCA	TGGAGGCAT	TGCCGGCGC	ACGGATCTTC	TGTGCATCGG
GCGCTGGTAC	CGGAAACCGT	ACCTCCGCA	ACGGCCCGCG	TGCTAGAAG	ACACGTAGCC
2590	2600	2610	2620	2630	2640
CGAAAATGGGC	ATCGGCAACA	CCACGATCGC	GGCCGGCATC	AATCTCGGCC	TTTATGGTGG
GCTTACCCCG	TAGCCGTTGT	GGTGCTAGGG	CCGGCGCTAG	TTAGAGCCGG	AAATACCAACC
2650	2660	2670	2680	2690	2700
CACGGCCGAA	GAATGGGTG	GTCCGGGTAC	CGGCTCGAG	GGCGAGGTGC	TGAAGCGCAA
GTGCCGGCTT	CTTACCCAGC	CAGGCCCCATG	GCCGAGGCTC	CCGCTCCACG	ACITCGCGTT
2710	2720	2730	2740	2750	2760
GATGCCCGCG	GTCGAAAAGG	CCGTGGCGCT	GCATCGCGAT	CACCTGTCCG	ATCCGCTCGA
CTAGCGGCGC	CAGCCTTTC	GGCACCGCGA	CGTAGCGCTA	GTGGACAGGC	TAGGCAGCT
2770	2780	2790	2800	2810	2820
ACTGATGCGT	CGCCTGGCG	GTCGTGAGAT	CGCGGCCATG	GCTGGCGCA	TCCTGGCCGC
TGACTACGCA	CGGGAGCCGC	CAGCACTCTA	GCGCCGGTAC	CGACCGCGT	AGGACCGGGCG
2830	2840	2850	2860	2870	2880
CCGCGTCCAG	AAGGTACCTG	TCATCATCGA	CGGCTACGTG	GCGACCGCTG	CGGCTTCGAT
GGCGCAGGTC	TTCCATGGAC	AGTAGTACCT	GCCGATGCAC	CGCTGGCGAC	GCCGAAGCTA

FIGURE 33.2

PL. 89/151

2890	2900	2910	2920	2930	2940
CCTGAAGGCG	GCCAAACCGT	CGGCCCTCGA	CCATTGCCCTG	ATCGGCCATG	TTTCGGGCAG
GGACTTCCGC	CGGTTGGGCA	GCCGGGAGGT	GGTAACGGAC	TAGCCGGTAC	AAAGCCCCGT
2950	2960	2970	2980	2990	3000
ACCGGGGCAT	CTGCGCGCGA	TCGAGAACGCT	CGGCAAGACG	CCGCTGCTGG	CACTCGGCAT
TGGCCCCGTA	GACGCGCGCT	AGCTCTTCGA	GCCGTTCTGC	GGCGACGACC	GTGACCGCTA
3010	3020	3030	3040	3050	3060
GC GGCTTGCG	GAAGGCACGG	GCGCGGCCCT	TGCCGCCGT	ATCGTCAAGG	CGGCGGCCG
CGCCGAACCG	CTTCCG-GCC	CGCGCCGGGA	ACGGCGGCCA	TAGCAGTTCC	GCCGCCGGCG
3070	3080	3090	3100	3110	3120
TTGCCACAGC	GCGATGGCGA	CTTTGCCCA	GGCCGGCGTC	AGCAACAAGG	AATAGTGAAG
AACGGTGTG	CCGTACCGCT	GGAAACGGGT	CCGGCCGCAG	TCGTTGTTCC	TTATCACTTC
3130	3140	3150	3160	3170	3180
TTCCGGCCGG	GCTTTGCAGG	AAGGCCGGCC	GGTTTCTGTC	CAAGGCCTGT	CACGGGCGCG
AAGGCCGGCC	CGAAACGTCC	TTCCGGCCGG	CCAAAGACAG	GTCCGGACA	GTGCCCGCGC
3190	3200	3210	3220	3230	3240
AAGCTGTCGC	GTGCCGGGC	TTGATGGATG	CGTCCTCTC	GCCTATCAA	AGCGCAAATG
TTCGACAGCG	CACGGCCCGG	AACTACCTAC	GCAGGAAGAG	CGGATAGGTT	TCGCGTTTAC
3250	3260	3270	3280	3290	3300
CGCGCCCTAG	CTATAGTCTT	GGGTGCCCTG	AACCGAGACC	GCCTTGCAATT	CGCCCTCAATC
CGCGGGGATC	GATATCAGAA	CCCACGGACG	ITGGCTCTGG	CGGAACGTAA	CGGGAGTTAG
3310	3320	3330	3340	3350	3360
ACGATGTCGA	AGCAAGCACA	GTTTCAGGCC	CTGTCGAGAC	GAATGGACG	CCAAGAACAC
TGCTACAGCT	TCGTTCGTGT	CAAAGTTGG	GACAGCTCTG	CTTACCTGC	GGTTCTTGTG
3370	3380	3390	3400	3410	3420
CACCGACCGC	ATTGGACAGA	CGGGCTCTGT	CGAGAACGAG	ACCGGCATTC	GGCATCTCTT
GTGGTGGCG	TAACCTGTCT	GCCCAAGGACA	GCTCTTCGTC	TGGCCGTAAG	CCGTAGAGAA
3430	3440	3450	3460	3470	3480
TGGCGCTGCG	AGCTATTGCG	TCGGCGGCC	CAAGCGGCTG	ATCGGGGAGG	CTGCTTTCG
ACGCCGACGC	TCGATAAGCG	AGCCGCCCG	GTTCGCCGAC	TAGCCGCTCC	GACGGAAAGC
3490	3500	3510	3520	3530	3540
CCACCGAGCTG	ATCGCCTTTC	CCGCGCGCAT	GATCGCTTTC	ATCATCGTCG	GCGCAACCTT
GGTGCCTCGAC	TAGCGGAAAC	GGCGGCCGTA	CTAGCGAAAG	TAGAGCAGC	CGCGTTGGAA
3550	3560	3570	3580	3590	3600
CTTCCAATAT	GTGGCGATGG	CGATCCTGTT	CCTGCTGATG	ATGGCCTTCG	AGGCAGATCAA
GAAGGTTATA	CACCGCTTAC	GCTAGGACAA	GGACGACTAC	TACCGGAAGC	TCCGCTAGTT
3610	3620	3630	3640	3650	3660
CACCGCAATC	GAGGAAATTG	TCGATCGCGT	TTCTCCCGAA	ATCTCGAAA	TGGGTAAGAA
GTGGCGTTAG	CTCCCTTAAAC	AGCTAGCGCA	AAGAGGGCTT	TAGAGCCTTT	ACCCATTCTT
3670	3680	3690	3700	3710	3720
CGCCAAGGAT	CTCGGCTCCT	TCGCCCTGCCT	CTGCGCTGATT	GTGCGCAACG	GTGTCTATGC
GCGGTTCTA	GAGCCGAGGA	AGCGGACGGA	GACGGACTAA	CAGCGGTGTC	CACAGATAAG
3730	3740	3750	3760	3770	3780
CGCCATATGTC	GTGATCTTCG	ACGGCTTCAT	GAAGTGACCG	GCTAGCGGGC	CGGCGCCCTC
GCGGATACAG	CACTAGAACG	TGCCGAAGTA	CTTGACTGGC	CGATCGCCCG	GCCGCGGAAAG
3790	3800	3810	3820	3830	3840
ACCCGATAAA	GCACATGGCG	ACGGAGCGGG	TTGCCCCCGG	GTACCGTGAC	GTGCGTCAA
TGGGCTATTT	CGTGTACGCC	AACGGGGGCC	CATGGCACTG	CAGCAGCTT	

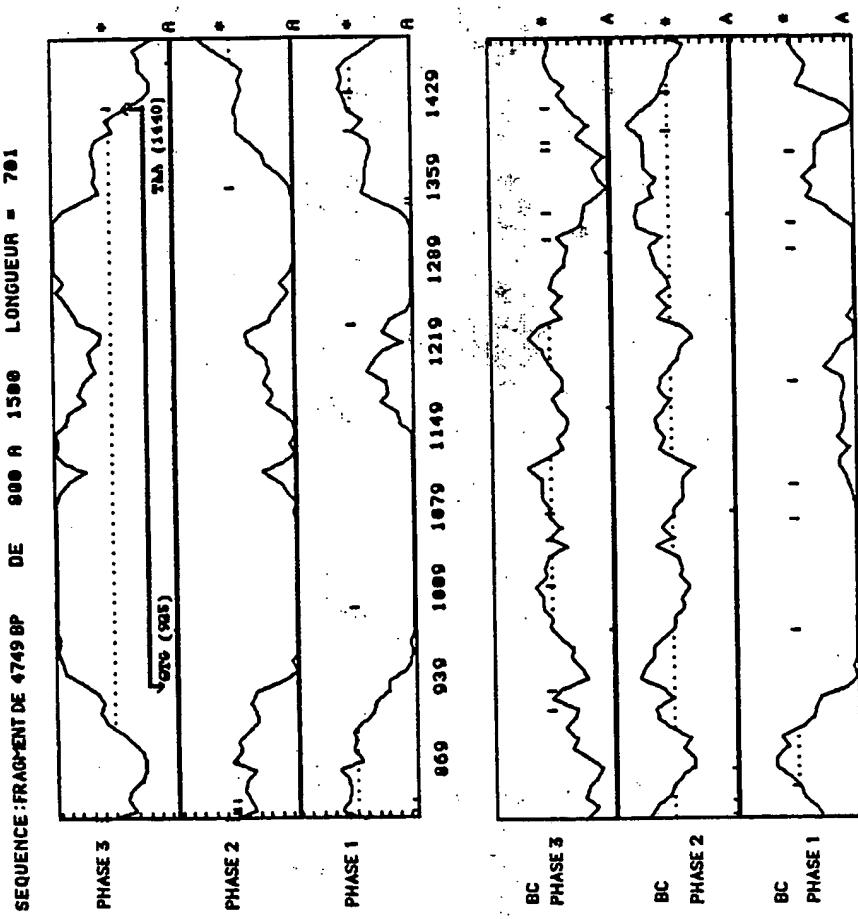
FIGURE 33.3

PL. 90/157  
3850      3860      3870      3880      3890      3900  
TCATCAGCCG GATCC  
AGTAGTCGGC CTAGG

FIGURE 33.4

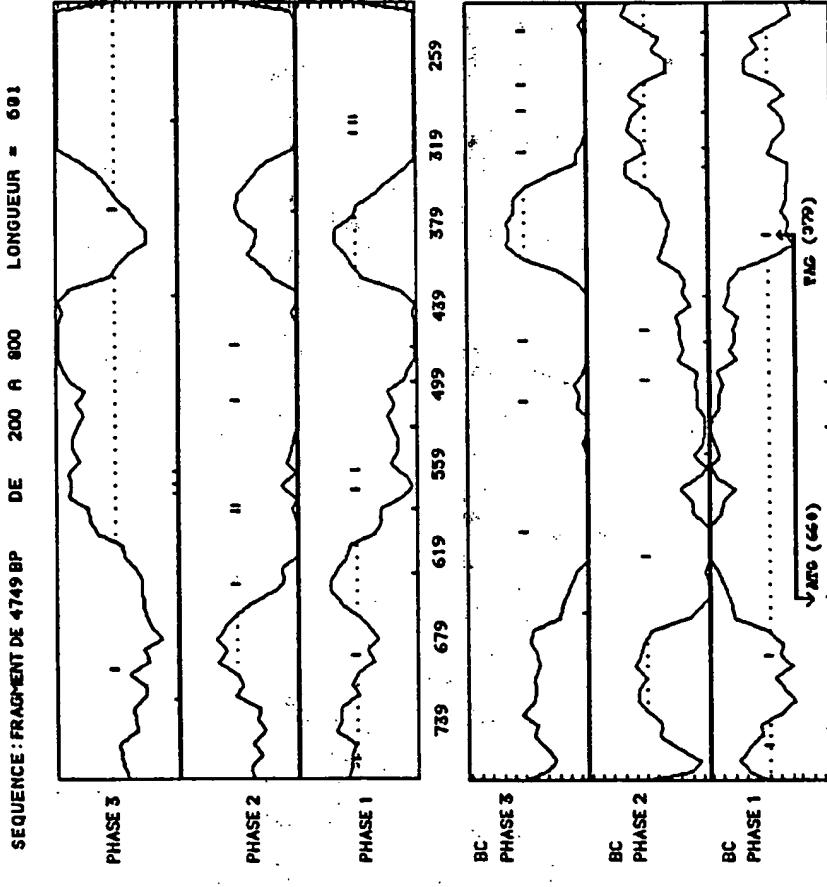
FEUILLE DE REMplacement

FIGURE 34



PHASE OUVERTE 15

A = ATG  
\* = STOP  
BC = BRIN COMPLEMENTAIRE



PHASE OUVERTE 14

A = ATG  
\* = STOP  
BC = BRIN COMPLEMENTAIRE

92 / 151

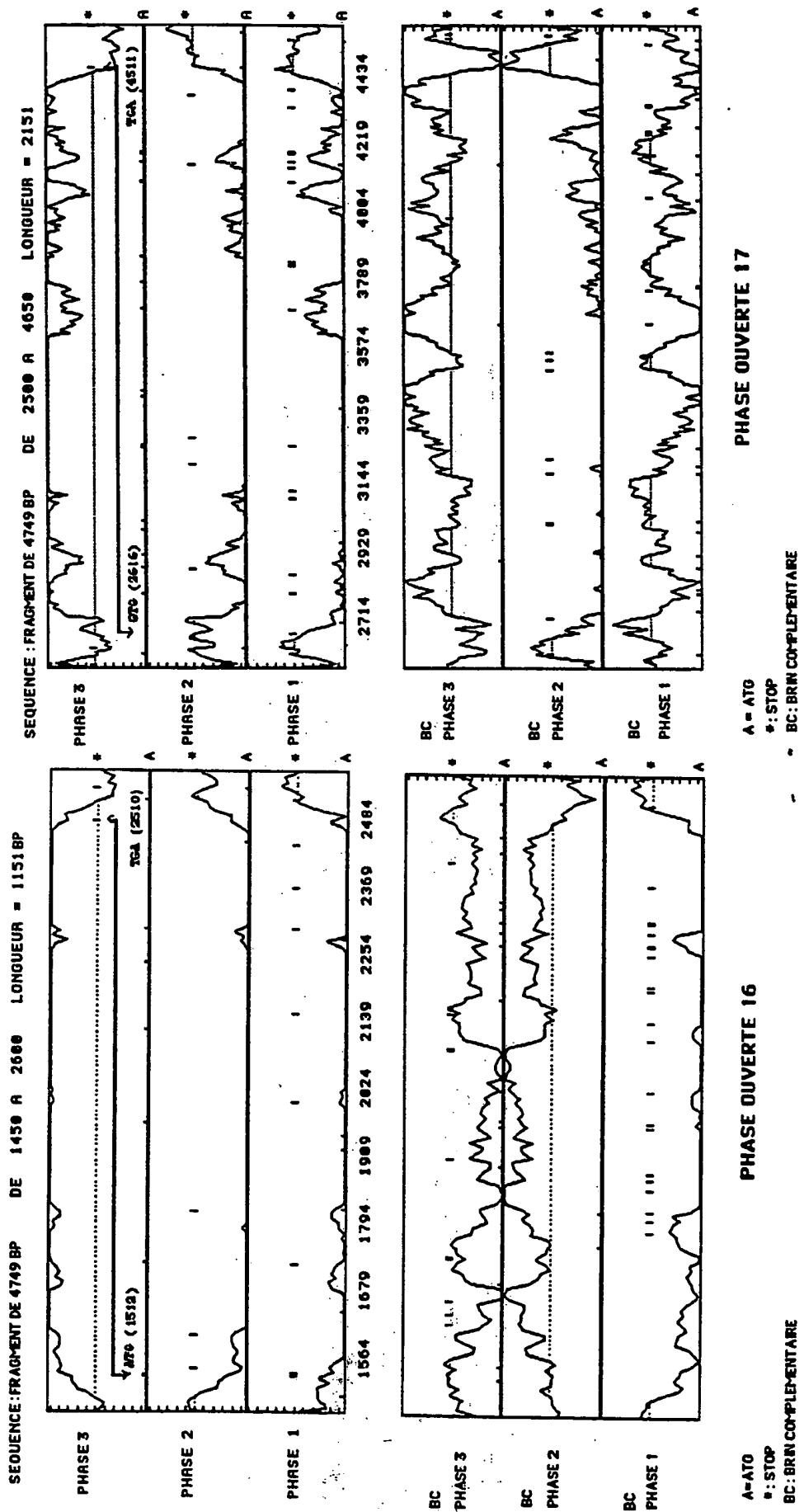


FIGURE 34.1

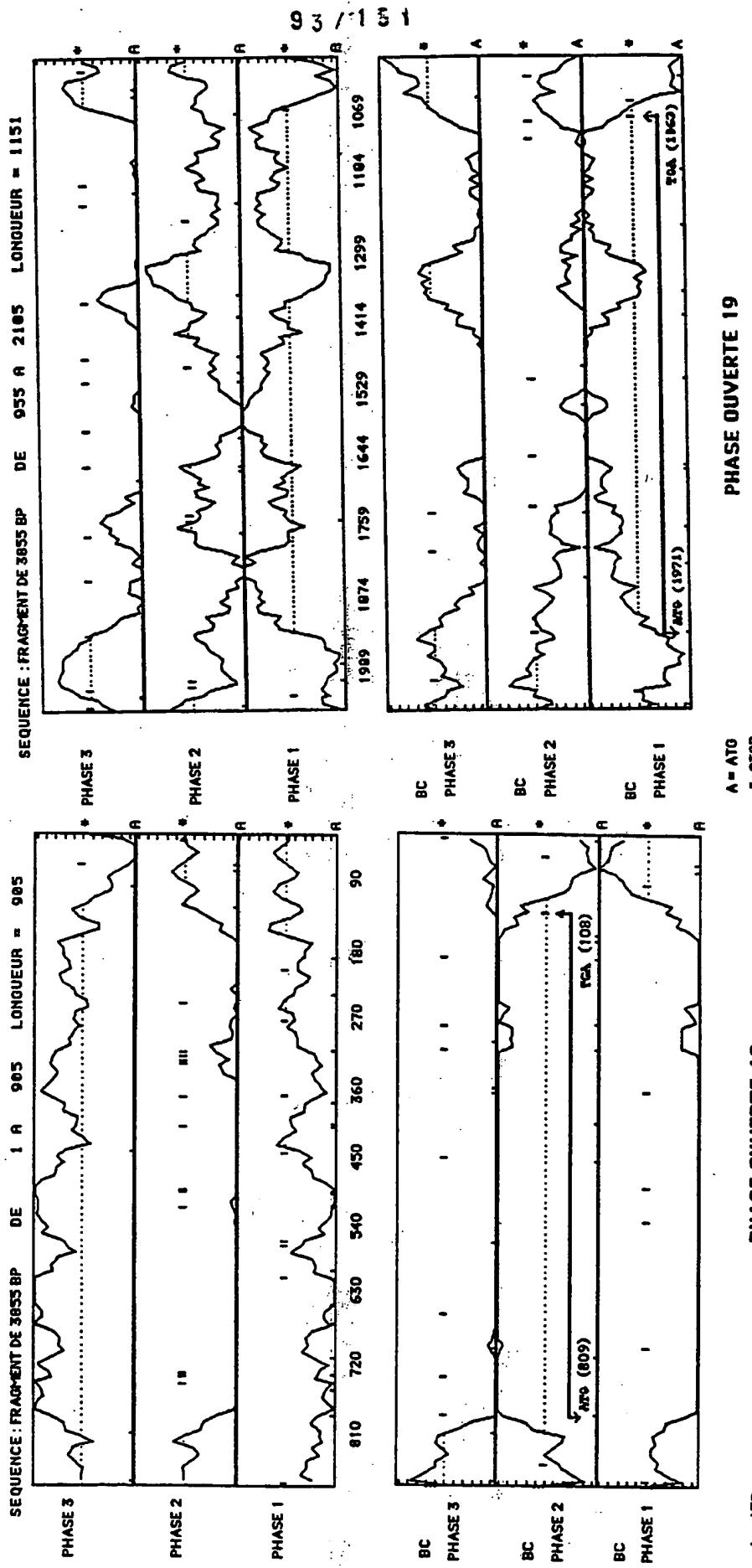


FIGURE 35

94 / 151

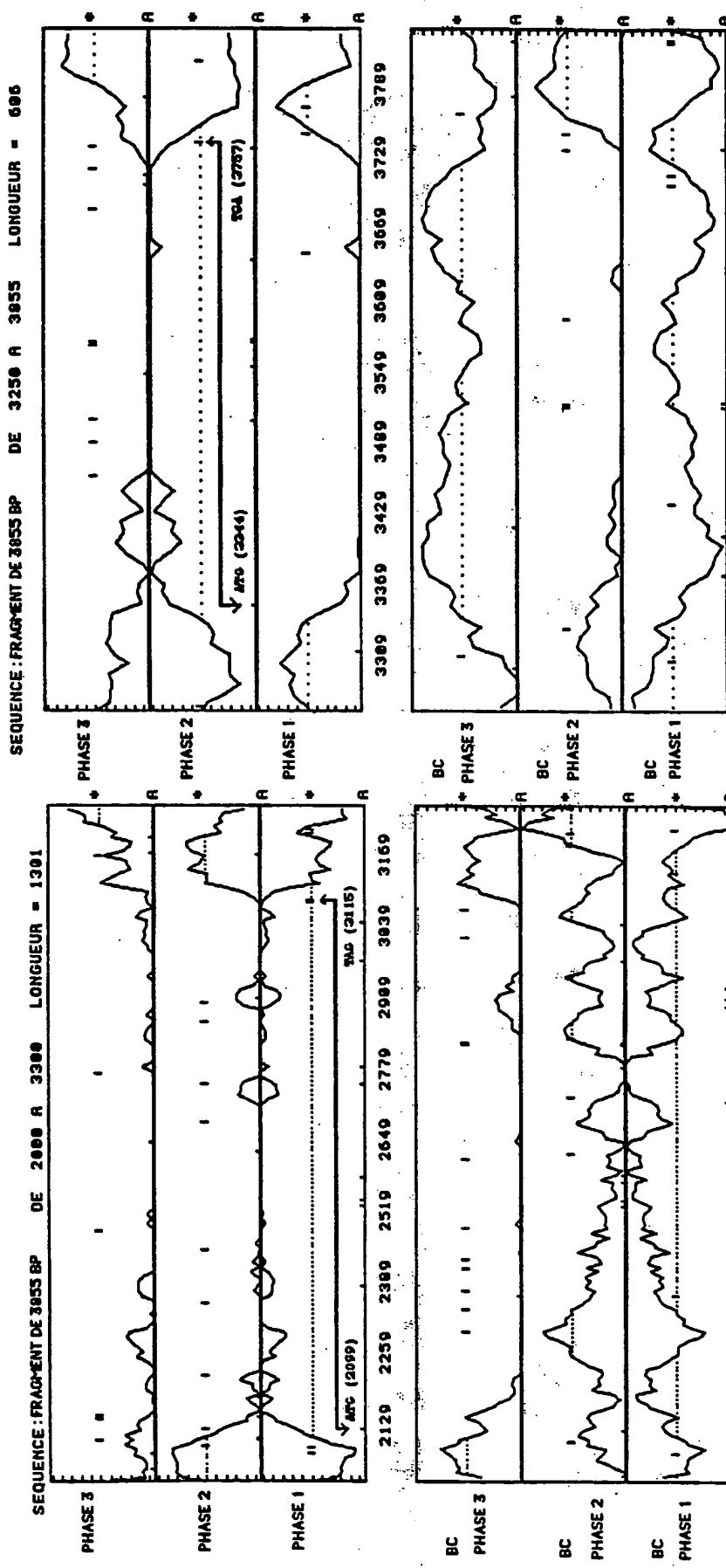


FIGURE 35.1

95 / 151

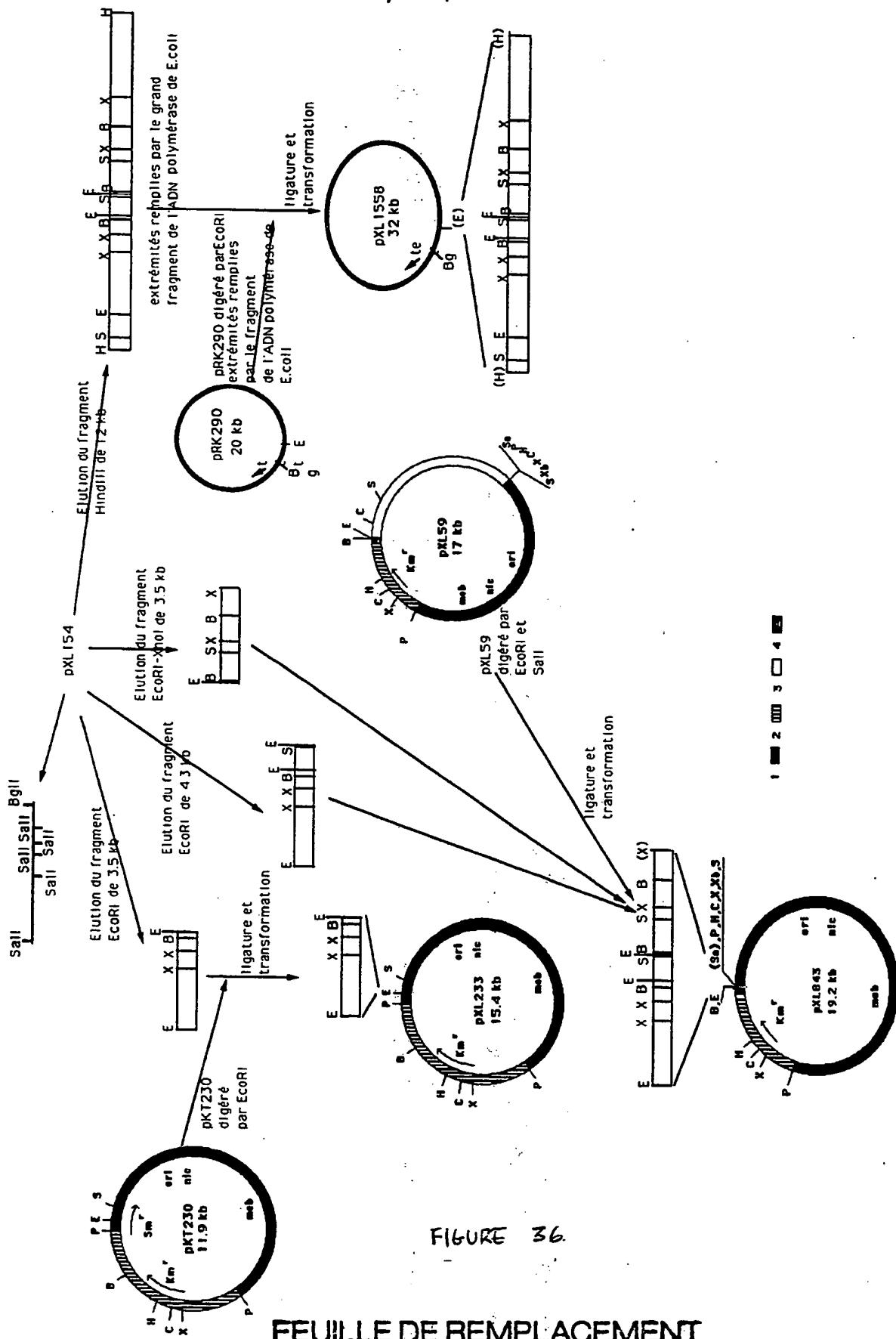


FIGURE 36.

FEUILLE DE REMPLACEMENT

96/151

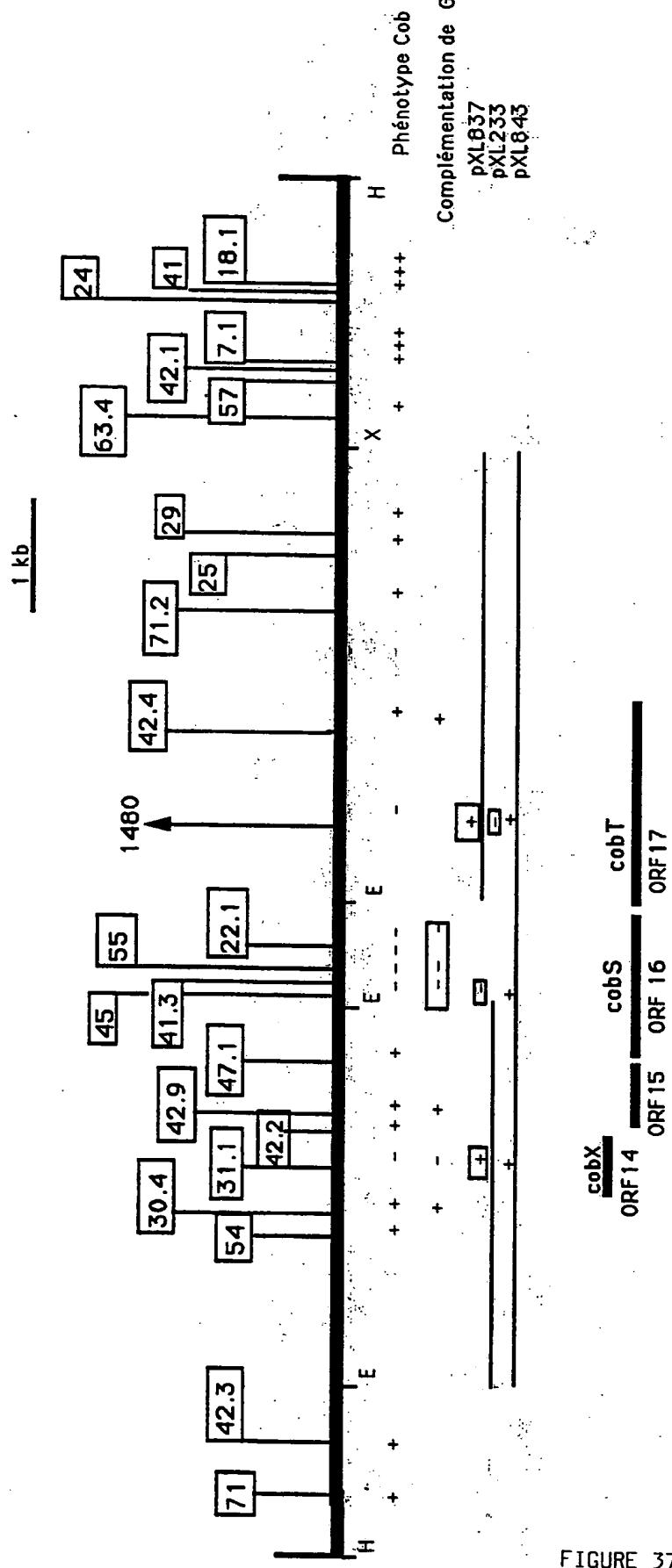


FIGURE 37

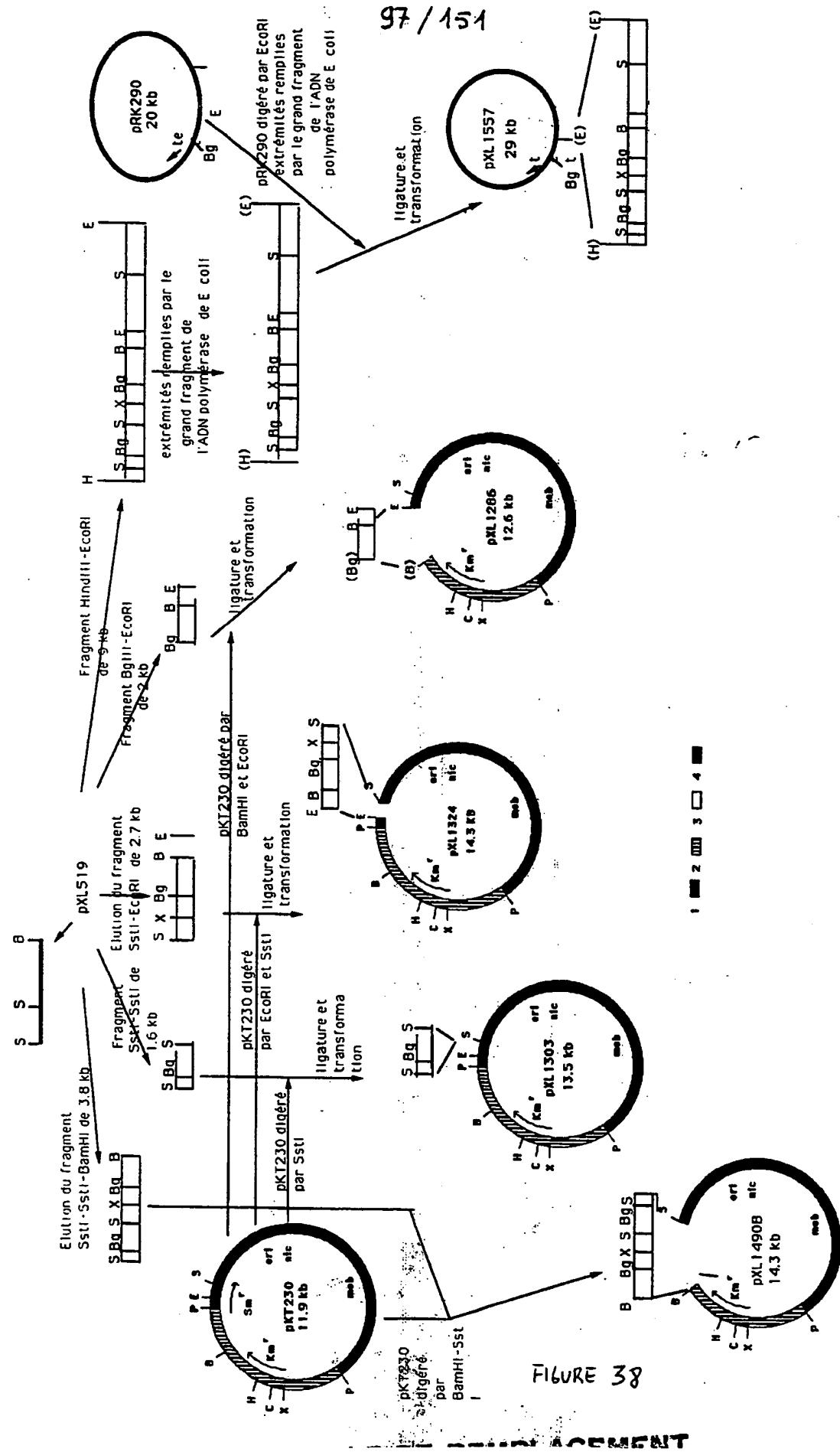
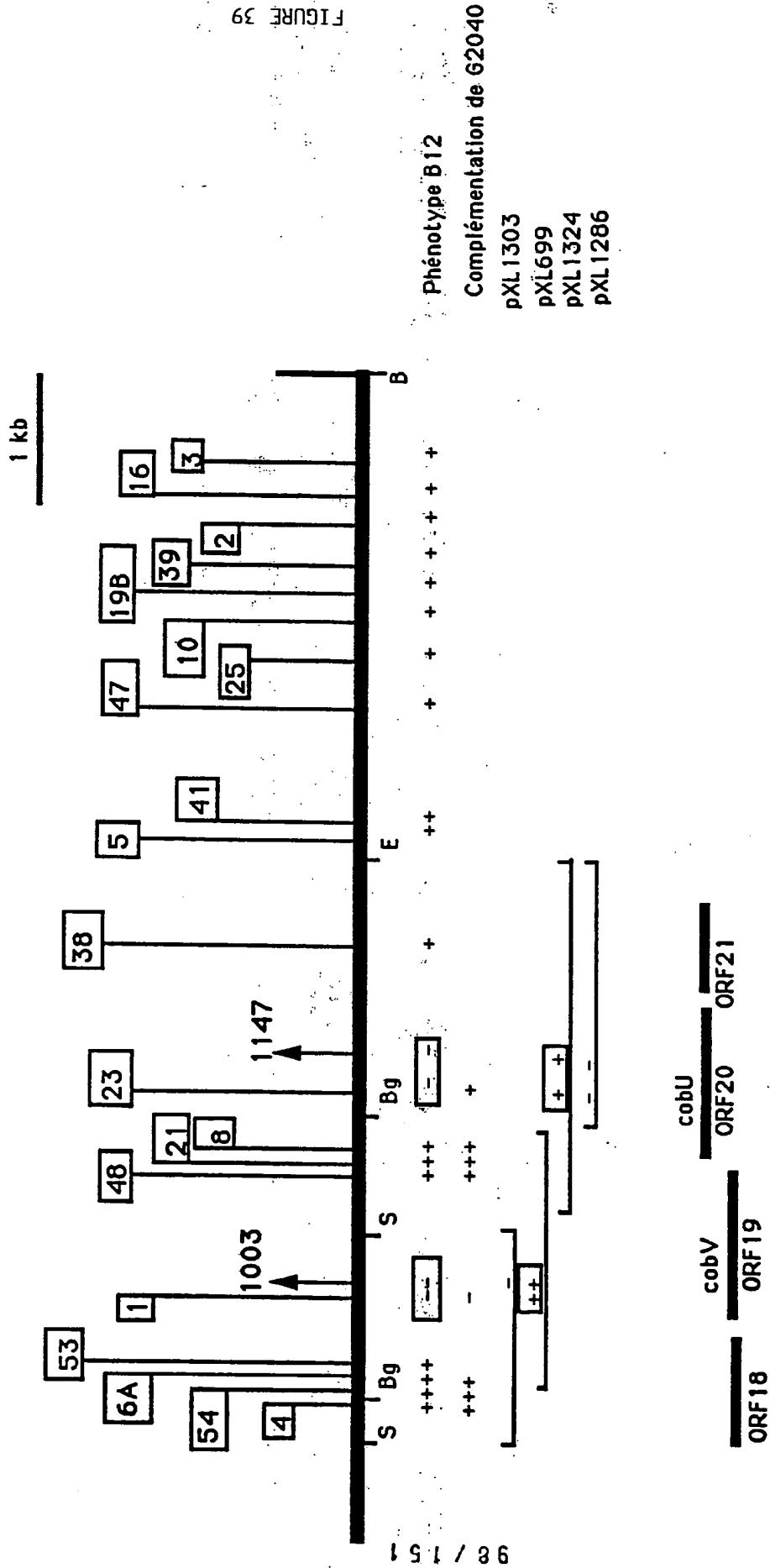


FIGURE 39



NOM = COBS

PL. 99/151  
PREMIER RESIDU = 1  
DERNIER RESIDU = 332

	NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	15	4.52
2	LEU	L	29	8.73
3	ILE	I	20	6.02
4	MET	M	7	2.11
5	VAL	V	33	9.94
6	SER	S	15	4.52
7	PRO	P	11	3.31
8	THR	T	24	7.23
9	ALA	A	32	9.64
10	TYR	Y	9	2.71
11	*	*	0	0.00
12	HIS	H	10	3.01
13	GLN	Q	10	3.01
14	ASN	N	12	3.61
15	LYS	K	13	3.92
16	ASP	D	28	8.43
17	GLU	E	15	4.52
18	CYS	C	3	0.90
19	TRP	W	4	1.20
20	ARG	R	22	6.63
21	GLY	G	20	6.02
22	-	-	0	0.00

RESIDUS = 332  
 Masse moleculaire (monoisotopique) = 36960.0000  
 Masse moleculaire (moyenne) = 36983.1797  
 INDEX DE POLARITE (%) = 44.88  
 POINT ISOELECTRIQUE = 6.34  
 DO 260 (1mg/ml) = 0.611 DO 280 (1mg/ml) = 0.891

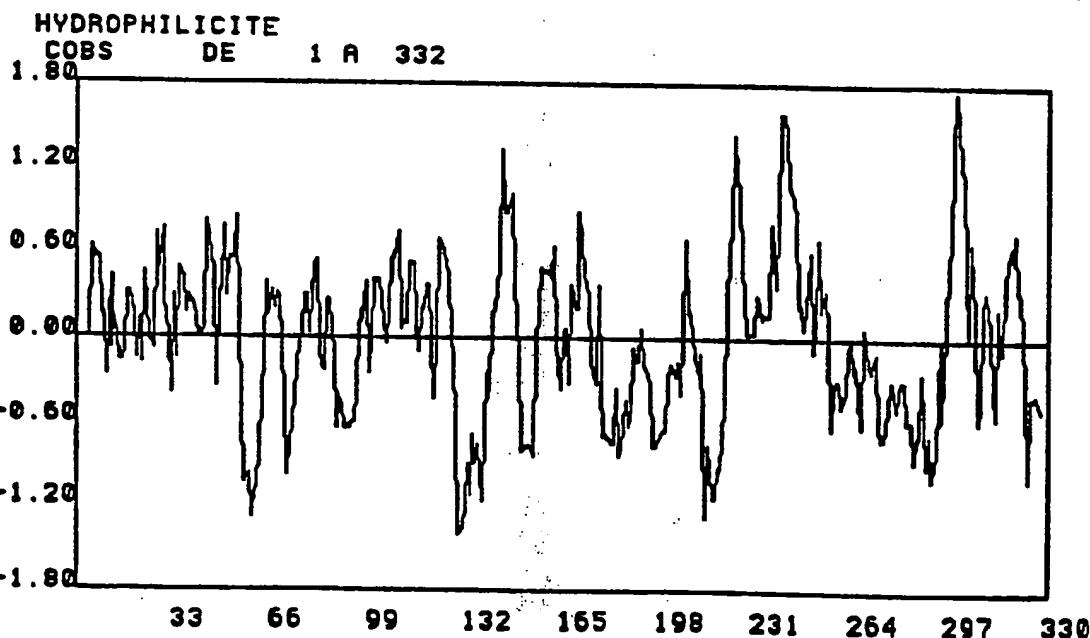


FIGURE 40

100/151

GENE cobs ET PROTEINE COBSSEQUENCE DU FRAGMENT Sali-Sali-Sali-Sali-Sali-Eg1 DE 4749 BP, DE 1512 A 2510

MetMetSerLysIleAspLeuAspIleSerAsnLeuProAspThrThrIleSerValArgGluValPheGlyIle  
 ATGATGAGCAAGATTGACCTCGACATTCACCTCCCGACACCACGATTCCGTCCGGAGGTTTCGGTATT  
 1521 1531 1541 1551 1561 1571 1581  
 AspThrAspLeuArgValProAlaTyrSerLysGlyAspAlaTyrValProAspLeuAspProAspTyrLeuPhe  
 GATACGGATTGCGCGTCCCTGCCATTTCGAAGGGCGACGCCATGTCCCAGTCTGGATCCGGACTACCTCTTC  
 1596 1606 1616 1626 1636 1646 1656  
 AspArgGluThrThrLeuAlaIleLeuAlaGlyPheAlaHisAsnArgArgValMetValSerGlyTyrHisGly  
 GACCGCGAAACGACCGCTGCCATTCTCGCAGGCTTCGCCAACACCGACGCGTGATGGTGTGCGCGTATCACGGC  
 1671 1681 1691 1701 1711 1721 1731  
 ThrGlyLysSerThrHisIleGluGlnValAlaAlaArgLeuAsnTrpProCysValArgValAsnLeuAspSer  
 ACCGGCAAGTCCACCCATATCGAGCAGGTCGCCGCCTCAACTGGCGTGCCTGCGTCAACCTCGATAGC  
 1746 1756 1766 1776 1786 1796 1806  
 HisValSerArgIleAspLeuValGlyLysAspAlaIleValValLysAspGlyLeuGlnValThrGluPheLys  
 CATGTCAGCCGTATCGACCTCGTCCGCAAGGACGCGATCGTCGTCAAGGACGCCCTGCAGTCACCGAATTCAAG  
 1821 1831 1841 1851 1861 1871 1881  
 AspGlyIleLeuProTrpAlaTyrGlnHisAsnValAlaLeuValPheAspGluTyrAspAlaGlyArgProAsp  
 GACGGCATCCTGCCCTGGGCTACCAGCACAAATGTCGGCTCGTGTTCGACGAAATACGATGCCGCCGGAC  
 1896 1906 1916 1926 1936 1946 1956  
 ValMetPheValIleGlnArgValLeuGluSerSerGlyArgLeuThrLeuAspGlnSerArgValIleArg  
 GTCATGTCGTATCCAGCGCTGCTGGATCCTCCGGCCCTGACGCTGCTCGACGAGGCCGTGTCATCCGT  
 1971 1981 1991 2001 2011 2021 2031  
 ProHisProAlaPheArgLeuPheAlaThrAlaAsnThrValGlyLeuGlyAspThrGlyLeuTyrHisGly  
 CCGCACCCGGCCTTCCGCCTGCGACACCGCTGGCCTCGGCACACGACCGCCCTATCACGGC  
 2046 2056 2066 2076 2086 2096 2106  
 ThrGlnGlnIleAsnGlnAlaGlnMetAspArgTrpSerIleValThrLeuAsnTyrLeuProHisAspLys  
 ACGCAGCAGATCAACCAGGCGCAGATGGACCGCTGGTCGATCGTCAACCACGCTGAACACTGCCGACGACAAG  
 2121 2131 2141 2151 2161 2171 2181  
 GluValAspIleValAlaAlaLysValLysGlyPheThrAlaAspLysGlyArgGluThrValSerLysMetVal  
 GAAAGTCGACATCGTCGCCGCCAGGTCAAGGGCTTACCGCCGACAAGGGCCGAGACCGTCTCCAAGATGGTA  
 2196 2206 2216 2226 2236 2246 2256  
 ArgValAlaAspLeuThrArgAlaAlaPheIleAsnGlyAspLeuSerThrValMetSerProArgThrValIle  
 CGTGTCCCGACCTCACGCGCCAGCCTCATCAATGGCAGTCTCGACTGTCATGAGCCCCTGACGGTCATC  
 2271 2281 2291 2301 2311 2321 2331  
 ThrTrpAlaGluAsnAlaHisIlePheGlyAspIleAlaPheArgValThrPheLeuAsnLysCysAsp  
 ACCTGGCCGAGAACGCCACATCTCGCGACATCGCTTCGCCCTCGCGTGACCTTCTCAACAAGTGCAC  
 2346 2356 2366 2376 2386 2396 2406  
 GluLeuGluArgAlaLeuValAlaGluHisTyrGlnArgAlaPheGlyIleGluLeuLysGluCysAlaAlaAsn  
 GAGCTGGAGCGGGCGCTGGTCGCCGAGCACTACCAGCGCCCTCGGCATCGAGCTGAAGGAATGCGCTGCCAAC  
 2421 2431 2441 2451 2461 2471 2481  
 IleValLeuGluAlaThrAla\*\*\*  
 ATCGTGCCTCGAAGCCACCGCCTGA  
 2496 2506

FIGURE 40.1

PL. 101 (15)

NOM = COBT

PREMIER RESIDU = 1  
DERNIER RESIDU = 631

	NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	16	2.54
2	LEU	L	56	8.87
3	ILE	I	29	4.60
4	MET	M	18	2.85
5	VAL	V	31	4.91
6	SER	S	33	5.23
7	PRO	P	24	3.80
8	THR	T	28	4.44
9	ALA	A	75	11.89
10	TYR	Y	8	1.27
11	*	*	0	0.00
12	HIS	H	7	1.11
13	GLN	Q	29	4.60
14	ASN	N	22	3.49
15	LYS	K	25	3.96
16	ASP	D	49	7.77
17	GLU	E	67	10.62
18	CYS	C	5	0.79
19	TRP	W	6	0.95
20	ARG	R	61	9.67
21	GLY	G	42	6.66
22	-	-	0	0.00

RESIDUS

= 631

Masse moleculaire (monoisotopique) = 70291.3984

Masse moleculaire (moyenne) = 70334.7656

INDEX DE POLARITE (%) = 50.87

POINT ISOELECTRIQUE = 5.10

DO 260 (1mg/ml) = 0.423 DO 280 (1mg/ml) = 0.610

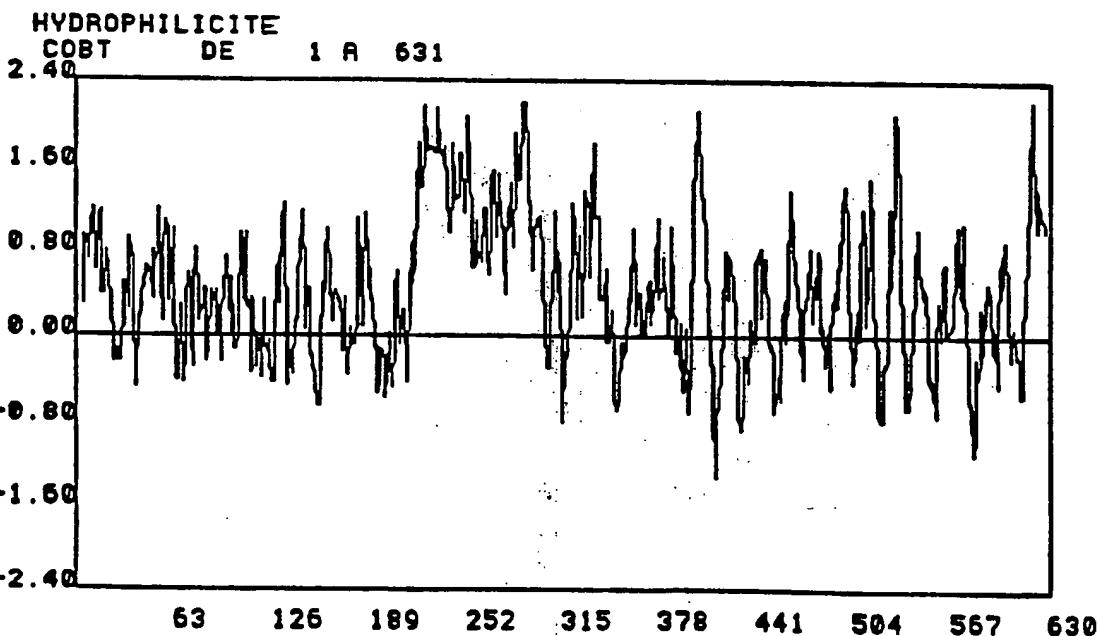


FIGURE 4D.2

102 / 151

GENE cobT ET PROTEINE COBTSEQUENCE DU FRAGMENT Sali-Sali-Sali-Sali-EgI DE 4749 BP, DE 2616 A 4511

Val Ser Ser Asn Ser Lys Ala Lys Pro Thr Thr Arg Glu Asn Ala Ala Glu Pro Phe Lys Arg Ala Leu Ser Gly  
 GTGAGCTCGAATTCAAGCAAAGCCAACCACGCGCAGAACATGCTGCGGAACCGTTCAAGCGGGCGCTTCGGC  
 2625 2635 2645 2655 2665 2675 2685  
 Cys Ile Arg Ser Ile Ala Gly Asp Ala Glu Val Glu Val Ala Phe Ala Asn Glu Arg Pro Gly Met Thr Gly Glu  
 TGCATCCGATCGATCGCGGGCGATGCCGAGGTGGAAGTCGCCTTCGCCAACGAGCGGCCGGCATGACCGGC  
 2700 2710 2720 2730 2740 2750 2760  
 Arg Ile Arg Leu Pro Glu Leu Ser Lys Arg Pro Thr Leu Glu Leu Ala Val Thr Arg Gly Leu Gly Asp Ser  
 CGCATCCGCTGCCGAACTTCCAAGCGCCGACCCCTGCAGGAACCTGCCGTGACCCGCCGGCTCGGTGACAGC  
 2775 2785 2795 2805 2815 2825 2835  
 Met Ala Leu Arg Lys Ala Cys Thr His Ala Arg Ile Gln Arg Thr Met Ser Pro Gln Gly Ala Asp Ala Arg Ala  
 ATGGCGCTGCGCAAGGGCTGTACGATGCCGGATCCAGCGCACCATGTCGCCAAGGGCGGACGCCCGCG  
 2850 2860 2870 2880 2890 2900 2910  
 Ile Phe Asp Ala Val Glu Gln Ala Arg Val Glu Ala Ile Gly Ser Leu Arg Met Ala Gly Val Ala Lys Asn Leu  
 ATCTTCGATGCCGGTGGAGCAGGCTCGTGTGAGGCCATGGCTCGTGCATGGCGGGTGTGCGCAAGAACCTC  
 2925 2935 2945 2955 2965 2975 2985  
 Asn Val Met Leu Glu Glu Lys Tyr Ala Lys Ala Asn Phe Ala Thr Ile Glu Arg Gln Ala Asp Ala Pro Leu Gly  
 AACGTCATGCTCGAAGAGAAATACGCCAACGGCGAACATTGCAACGATCGAGGCCAGGCCGACGCCCGCTCGC  
 3000 3010 3020 3030 3040 3050 3060  
 Glu Ala Val Ala Leu Leu Val Arg Glu Lys Leu Thr Gly Gln Lys Pro Pro Ala Ser Ala Gly Lys Val Leu Asp  
 GAGGCCGTAGCGCTGCTGGTGCGCGAGAACGCTGACGGGCCAGAACGCCGCCGCTGCCGGCAAGGTGCTCGAC  
 3075 3085 3095 3105 3115 3125 3135  
 Leu Trp Arg Glu Phe Ile Glu Gly Lys Ala Ala Gly Asp Ile Glu His Leu Ser Ser Thr Ile Asn Asn Gln Gln  
 CTCTGGCGCAGTTCATCGAGGGCAAGGCTGCCGGCACATTGAGCACCTGTCGTGACGATCAACAACCAGCAG  
 3150 3160 3170 3180 3190 3200 3210  
 Ala Phe Ala Arg Val Val Arg Asp Met Leu Thr Ser Met Glu Val Ala Glu Lys Tyr Gly Asp Asp Asp Asn Glu  
 GCCTTGCCCCGGTCGTTCGCACATGCTGACCTCGATGGAAGTCGCCAGAAATACGGTGACGACGACAACGAG  
 3225 3235 3245 3255 3265 3275 3285  
 Pro Asp Glu Gln Glu Ser Glu Thr Asp Glu Asp Gln Pro Arg Ser Gln Glu Gln Asp Glu Asn Ala Ser Asp Glu  
 CCGGACGAGCAGGAAAGCGAGACGACAGGCCGCGCAGGCCAGGAGCAGGACGAGAACGCCAGCGACGAG  
 3300 3310 3320 3330 3340 3350 3360  
 Glu Ala Gly Asp Asp Ala Ala Pro Ala Asp Glu Asn Gln Ala Ala Glu Glu Gln Met Glu Glu Gly Glu Met Asp  
 GAAGCCGGCGACGATGCCGCACCCGCCACGAGAACCAAGGCTGCCAACAGCAGATGGAAGAACGGCGAGATGGAC  
 3375 3385 3395 3405 3415 3425 3435  
 Gly Ala Glu Ile Ser Asp Asp Asp Leu Gln Asp Glu Gly Asp Glu Asp Ser Glu Thr Pro Gly Glu Val Lys Arg  
 GGCGCGGAGATCTCCGACGACGATCTCCAGGACGAAGGCCAGGCCAGGCCAGGAGCAGCAGAACGCCGGCAGGTCAAGCGT  
 3450 3460 3470 3480 3490 3500 3510  
 Pro Asn Gln Pro Phe Ala Asp Phe Asn Glu Lys Val Asp Tyr Ala Val Phe Thr Arg Glu Phe Asp Glu Thr Ile  
 CCGAACCCAGCCCTTCGCCGACTTCAACGAGAACGGTCACTACGCCGTCTTCACCCCGCAGGTTCGACGAGACGATT  
 3525 3535 3545 3555 3565 3575 3585  
 Ala Ser Glu Glu Leu Cys Asp Glu Ala Glu Leu Asp Arg Leu Arg Ala Phe Leu Asp Lys Gln Leu Ala His Leu  
 GCCTCGGAAGAGCTTGCACGAGGCCGAGCTCGACCGGCTGCCGCCTTCCTCGACAAGCAGCTGCCCATCTT  
 3600 3610 3620 3630 3640 3650 3660  
 Gln Gly Ala Val Gly Arg Leu Ala Asn Arg Leu Gln Arg Arg Leu Met Ala Gln Gln Asn Arg Ser Trp Glu Phe  
 CAAGGCCGGTCGGCCGCTTGCCAACCGCCTGCAACGCCGCTGCAGCGCCGCTGATGGCGCAGCAGAACCGCTCTGGAGTT  
 3675 3685 3695 3705 3715 3725 3735  
 Asp Leu Glu Glu Gly Tyr Leu Asp Ser Ala Arg Leu Gln Arg Ile Ile Ile Asp Pro Met Gln Pro Leu Ser Phe  
 GATCTCGAACAGAGGGTATCTCGATTGGCGCGCTTCAGCGCATCATCGATCCGATGCGAGCCGCTTCCCTTC  
 3750 3760 3770 3780 3790 3800 3810  
 Lys Arg Glu Lys Asp Thr Asn Phe Arg Asp Thr Val Val Thr Leu Leu Ile Asp Asn Ser Gly Ser Met Arg Gly  
 AACGCGAAAAGGACACCAACTCCGCGATACCGTCGTGACGCTGCTGATCGACAATTCCGGCTCGATGCGCGGC  
 3825 3835 3845 3855 3865 3875 3885  
 Arg Pro Ile Thr Val Ala Ala Thr Cys Ala Asp Ile Leu Ala Arg Thr Leu Glu Arg Cys Gly Val Lys Val Glu  
 CGTCCGATCACGGTTGCCGCCACCTGCGCCGATATCCTGGCGCGACGCTCGAGCGCTGCCGTCAAGGTCGAG  
 3900 3910 3920 3930 3940 3950 3960  
 Ile Leu Gly Phe Thr Thr Lys Ala Trp Lys Gly Gly Gln Ser Arg Glu Lys Trp Leu Ala Gly Gly Lys Pro Gln  
 ATCCCTGGTTTACCAAGGCGTGGAAAGGGTGGCGAGTCACGCCAGAAGTGGCTGCCGGCGCAAGCCACAG  
 3975 3985 3995 4005 4015 4025 4035

FIGURE 40.3

*103/151*

AlaProGlyArgLeuAsnAspLeuArgHisIleValTyrLysSerAlaAspAlaProTrpArgArgAlaArgArg  
GCCCGGGTCGCCTCAACGACCTGCGACACATCGTCTACAAGTCTGCCGACGCTCCGTGGCGCCGGCACGACG  
4050 4060 4070 4080 4090 4100 4110  
AsnLeuGlyLeuMetMetArgGluGlyLeuLeuLysGluAsnIleAspGlyGluAlaLeuIleTrpAlaHisGlu  
AATCTGGCCTGATGATGCGGGAAAGGCCTGCTCAAGGAAAACATCGACGGCGAGGCCTTGATTGGCGCATGAG  
4125 4135 4145 4155 4165 4175 4185  
ArgLeuMetAlaArgArgGluGlnArgArgIleLeuMetMetIleSerAspGlyAlaProValAspAspSerThr  
CGGCTGATGGCGCGCGAACAGCGGCCATCCTGATGATGATTGGACGGCGCCGGTCGACGACTCGACG  
4200 4210 4220 4230 4240 4250 4260  
LeuSerValAsnProGlyAsnTyrLeuGluArgHisLeuArgAlaValIleGluGlnIleGluThrArgSerPro  
CTGTCGGTCAATCCAGGAAACTATCTGGAGCGTCACCTGCGCGCGTCATCGAGCAGATCGAAACGCGCTCGCCG  
4275 4285 4295 4305 4315 4325 4335  
ValGluLeuLeuAlaIleGlyIleGlyHisAspValThrArgTyrTyrArgArgAlaValThrIleValAspAla  
GTGGAACTGCTGGCGATCGGTATCGGCCACGACGTGACGCGCTACTATCGCCGTGCCGTACCATCGTCGATGCC  
4350 4360 4370 4380 4390 4400 4410  
AspGluLeuAlaGlyAlaMetThrGluGlnLeuAlaAlaLeuPheGluAspGluSerGlnArgArgGlySerSer  
GATGAGCTTGCCGGCGCGATGACCGAACAGCTGGCCGACTCTCGAGGACGAAAGCCAGCGCCGCGTTCTTCG  
4425 4435 4445 4455 4465 4475 4485  
ArgLeuArgArgAlaGly\*\*\*  
CGTCTTCGCCGCGCCGGGTGA  
4500 4510

FIGURE 40.4

PL. 104/151

NOM = COBX

PREMIER RESIDU = 1  
DERNIER RESIDU = 93

	NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	3	3.23
2	LEU	L	8	8.60
3	ILE	I	5	5.38
4	MET	M	3	3.23
5	VAL	V	5	5.38
6	SER	S	6	6.45
7	PRO	P	6	6.45
8	THR	T	5	5.38
9	ALA	A	10	10.75
10	TYR	Y	0	0.00
11	*	*	0	0.00
12	HIS	H	7	7.53
13	GLN	Q	2	2.15
14	ASN	N	2	2.15
15	LYS	K	3	3.23
16	ASP	D	3	3.23
17	GLU	E	10	10.75
18	CYS	C	0	0.00
19	TRP	W	1	1.08
20	ARG	R	7	7.53
21	GLY	G	7	7.53
22	-	-	0	0.00

RESIDUS = 93

Masse moleculaire (monoisotopique) = 10279.2354

Masse moleculaire (moyenne) = 10285.6309

INDEX DE POLARITE (%) = 48.39

POINT ISOELECTRIQUE = 6.94

DO 260 (1mg/ml) = 0.411 DO 280 (1mg/ml) = 0.541

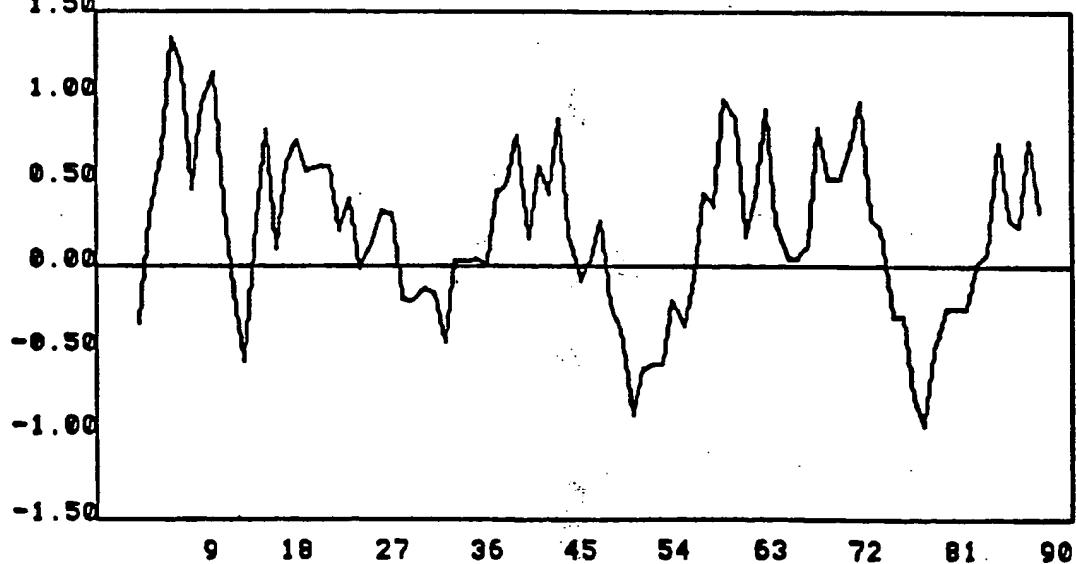
HYDROPHILICITE  
COBX DE 1 A 93

FIGURE 40.5

105/151

GENE cobX ET PROTEINE COBXSEQUENCE DU FRAGMENT BglI-SalI-SalI-SalI-SalI DE 4749 BP, DE 4089 A 4370

MetSerLeuThrGluThrIleGluLysLeuIleGluAlaPheHisProGluArgLeuGluValIleAsnGlu  
ATGTCGCTCACCGAGACCATCGAAAGAAGCTGATCGAGGCCCTCCACCCCTGAACGGCTCGAGGTCAACGAG  
4098 4108 4118 4128 4138 4148 4158  
SerHisGlnHisThrGlyHisGlnProGlyPheAspGlyThrGlyGluSerHisMetArgValArgIleValSer  
AGCCATCAGCATACCGGCCATCAGCCGGGTTCGATGGTACCGGGCAGTCCCACATGCGGGTGCGTATCGTTCT  
4173 4183 4193 4203 4213 4223 4233  
SerAlaPheAlaGlyMetSerArgValAlaArgHisArgAlaIleAsnAspLeuLeuLysProGluLeuAspAla  
AGCGCCTTGCCGGCATGAGCCGTGTCGCCGCCACCGCGCCATCAATGATCTCCTGAAGCCAGAACTCGACGCC  
4248 4258 4268 4278 4288 4398 4308  
GlyLeuHisAlaLeuAlaValGluProAlaAlaProGlyGluProThrArgTrp\*\*  
GGCCTGCATGCGCTCGCCGTCGAGCCGGCAGCCCCCGGCGAGCCGACCCGCTGGTAG  
4323 4333 4343 4353 4363

FIGURE 40.6

PL. 106 (15)

NOM = COBU

PREMIER RESIDU = 1  
DERNIER RESIDU = 338

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	9	2.66	1323.62	3.82
2	LEU	L	36	10.65	4071.03	11.75
3	ILE	I	19	5.62	2148.60	6.20
4	MET	M	9	2.66	1179.36	3.40
5	VAL	V	21	6.21	2080.44	6.00
6	SER	S	13	3.85	1131.42	3.26
7	PRO	P	16	4.73	1552.84	4.48
8	THR	T	17	5.03	1717.81	4.96
9	ALA	A	61	18.05	4333.26	12.50
10	TYR	Y	3	0.89	489.19	1.41
11	*	*	0	0.00	0.00	0.00
12	HIS	H	8	2.37	1096.47	3.16
13	GLN	Q	6	1.78	768.35	2.22
14	ASN	N	9	2.66	1026.39	2.96
15	LYS	K	12	3.55	1537.14	4.43
16	ASP	D	13	3.85	1495.35	4.31
17	GLU	E	22	6.51	2838.94	8.19
18	CYS	C	5	1.48	515.05	1.49
19	TRP	W	3	0.89	558.24	1.61
20	ARG	R	16	4.73	2497.62	7.21
21	GLY	G	40	11.83	2280.86	6.58
22	-	-	0	0.00	0.00	0.00

RESIDUS = 338

Masse moleculaire (monoisotopique) = 34659.9844

Masse moleculaire (moyenne) = 34681.9609

INDEX DE POLARITE (%) = 34.32

POINT ISOELECTRIQUE = 6.21

DO 260 (1mg/ml) = 0.416 DO 280 (1mg/ml) = 0.584

## HYDROPHILICITE

COBU DE 1 A 338

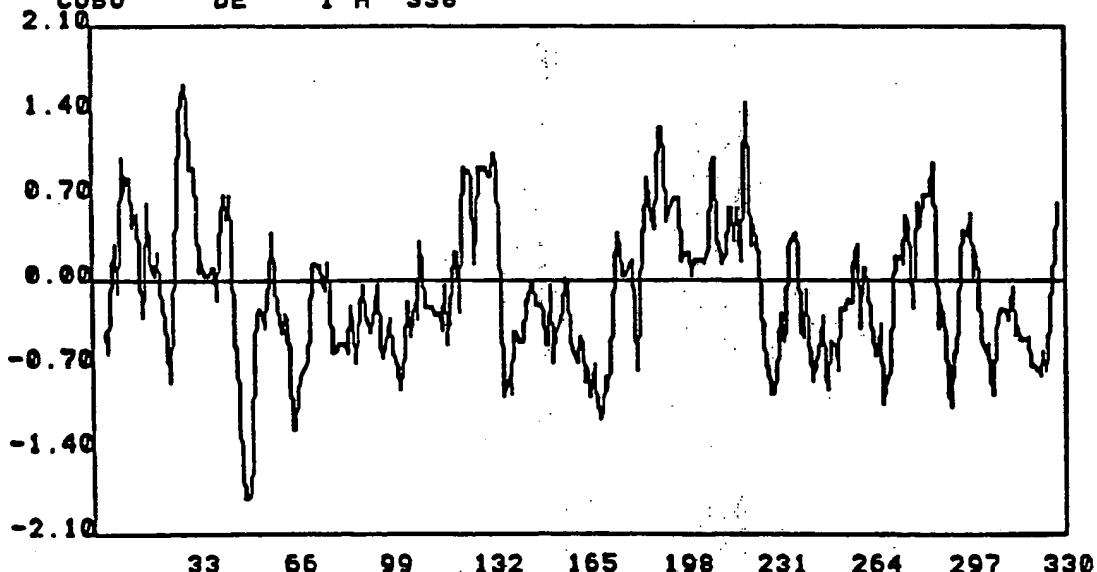


FIGURE 41

107 / 151

GENE cobU ET PROTEINE COBUSEQUENCE DU FRAGMENT SstI-SstI-BamHI DE 3855 BP, DE 2099 A 3115

MetSerAlaSerGlyLeuProPheAspAspPheArgGluLeuLeuArgAsnLeuProGlyProAspAlaAlaAla  
 ATGAGTGCCAGCGGGCTGCCGTTGATGATTTCGCAATTGTTGCGAACCTGCCGGCCGGATGCGGCAGCC  
     2108      2118      2128      2138      2148      2158      2168  
 LeuValAlaAlaArgGluArgAspAlaGlnLeuThrLysProProGlyAlaLeuGlyArgLeuGluGluIleAla  
 CTCGTTGCCGCGGGAGCAGCCCAGCTGACGAAGCCGCCGGCGCCTCGGCCGCTCGAGGAAATCGCC  
     2183      2193      2203      2213      2223      2233      2243  
 PheTrpLeuAlaAlaTrpThrGlyLysAlaProValValAsnArgProLeuValAlaIlePheAlaGlyAsnHis  
 TTCTGGCTCGCCGCTGGACGGCAAGGGCCGGTGGTCAACCGGCCGCTGGTGGCGATCTTGCCGCAACCAC  
     2258      2268      2278      2288      2298      2308      2318  
 GlyValThrArgGlnGlyValThrProPheProSerSerValThrAlaGlnMetValGluAsnPheAlaAlaGly  
 GGCCTCACCCGCCAGGGGTGACCCCGTCCCGTCATCCGTACCCGACAGATGGTCGAGAATTTCGCCGCCGGT  
     2333      2343      2353      2363      2373      2383      2393  
 GlyAlaAlaIleAsnGlnIleCysValSerHisAspLeuGlyLeuLysValPheAspLeuAlaLeuGluTyrPro  
 GGCGCTGCGATCAACCAGATCTGCGTCAGCCACGACCTCGGGCTGAAAGGTCTCGACACTCGAATACCCG  
     2408      2418      2428      2438      2448      2458      2468  
 ThrGlyAspIleThrGluGluAlaAlaLeuSerGluArgAspCysAlaAlaThrMetAlaPheGlyMetGluAla  
 ACCGGTGATATCACCGAGGAAGCCGCGCTGTCGGAGCGATTGCGCCGCGACCATGGCTTGGCATGGAGGCG  
     2483      2493      2503      2513      2523      2533      2543  
 IleAlaGlyGlyThrAspLeuLeuCysIleGluMetGlyIleGlyAsnThrThrIleAlaAlaAlaIleAsn  
 ATTGCCGGCGGACGGATCTCTGTGCATCGGCAAATGGCATCGGCAACACCACGATCGCGGCCGCGATCAAT  
     2558      2568      2578      2588      2598      2608      2618  
 LeuGlyLeuTyrGlyGlyThrAlaGluGluTrpValGlyProGlyThrGlySerGluGlyGluValLeuLysArg  
 CTCGGCCTTATGGTGGCACGGCGAAGAATGGGTGGCTCCGGTACCGGCTCCGAGGGCGAGGTGCTGAAGCGC  
     2633      2643      2653      2663      2673      2683      2693  
 LysIleAlaAlaValGluLysAlaValAlaLeuHisArgAspHisLeuSerAspProLeuGluLeuMetArgArg  
 AAGATCGCCGCGGTGCAAAAGGCCGTGGCGTCGATCACCTGTCGATCCGCTCGAACTGATGCGTCGC  
     2708      2718      2728      2738      2748      2758      2768  
 LeuGlyGlyArgGluIleAlaAlaMetAlaGlyAlaIleLeuAlaAlaArgValGlnLysValProValIleIle  
 CTCGGCGGTGAGATCGCGGCCATGGCTGGCGCCATGCTGGCCGCCCTCGAGAAGGTACCTGTCATCATC  
     2783      2793      2803      2813      2823      2833      2843  
 AspGlyTyrValAlaThrAlaAlaAlaSerIleLeuLysAlaAlaAsnProSerAlaLeuAspHisCysLeuIle  
 GACGGCTACGTGGCGACCGCTGCGCTTCGATCCTGAAGCGGCCAACCCGTCGGCCCTCGACCATTCGCTGATC  
     2858      2868      2878      2888      2898      2908      2918  
 GlyHisValSerGlyGluProGlyHisLeuArgAlaIleGluLysLeuGlyLysThrProLeuLeuAlaLeuGly  
 GCCCATGTTGGCGAACCGGGGCATCTGGCGCGATCGAGAACGCTGGCAAGACGCCGCTGGCACTCGGC  
     2933      2943      2953      2963      2973      2983      2993  
 MetArgLeuGlyGluGlyThrGlyAlaAlaLeuAlaAlaGlyIleValLysAlaAlaAlaCysHisSerGly  
 ATGCGGCTTGGCGAACGGCACGGCGCGGCCATTGCCGCCGGTATCGTCAAGGCAGGCCGCTTGCCACAGCGGC  
     3008      3018      3028      3038      3048      3058      3068  
 MetAlaThrPheAlaGlnAlaGlyValSerAsnLysGly\*\*\*\*  
 ATGGCGACCTTGCCCCAGGCCGGCGTCAGCAACAAGGAATAG  
     3083      3093      3103      3113

FIGURE 41.2

NOM = COBV

PREMIER RESIDU = 1  
DERNIER RESIDU = 302

PL. 108 / 15 /

	NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	18	5.96
2	LEU	L	39	12.91
3	ILE	I	13	4.30
4	MET	M	10	3.31
5	VAL	V	23	7.62
6	SER	S	18	5.96
7	PRO	P	12	3.97
8	THR	T	10	3.31
9	ALA	A	63	20.86
10	TYR	Y	3	0.99
11	*	*	0	0.00
12	HIS	H	3	0.99
13	GLN	Q	6	1.99
14	ASN	N	2	0.66
15	LYS	K	5	1.66
16	ASP	D	10	3.31
17	GLU	E	7	2.32
18	CYS	C	3	0.99
19	TRP	W	2	0.66
20	ARG	R	19	6.29
21	GLY	G	36	11.92
22	-	-	0	0.00

RESIDUS = 302  
 Masse moleculaire (monoisotopique) = 30642.3359  
 Masse moleculaire (moyenne) = 30662.0820  
 INDEX DE POLARITE (%) = 26.49  
 POINT ISOELECTRIQUE = 9.97  
 DO 260 (1mg/ml) = 0.391 DO 280 (1mg/ml) = 0.479

## HYDROPHILICITE

COBU DE 1 R 302

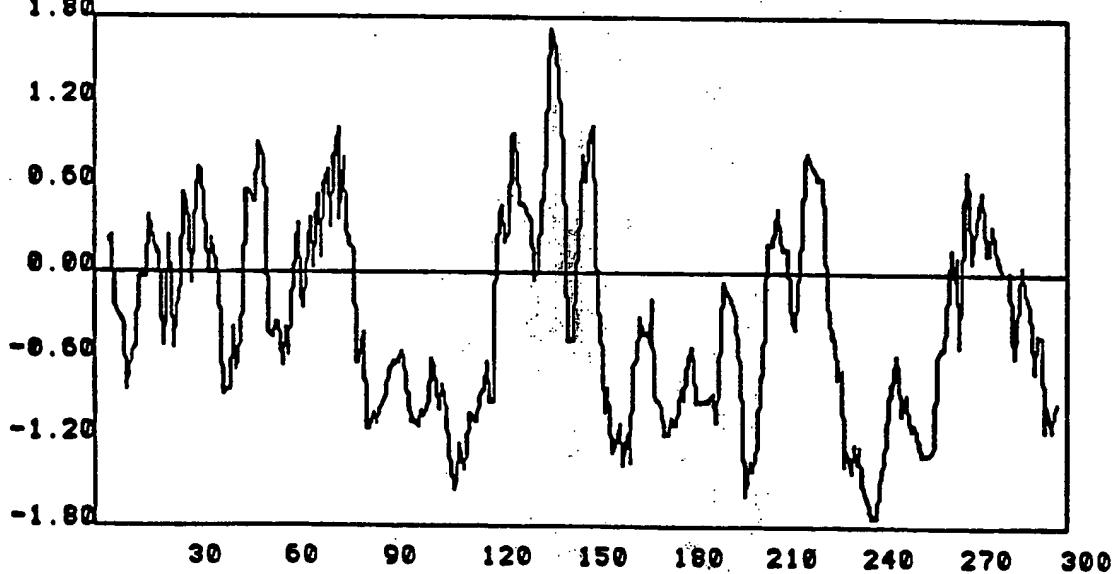


FIGURE 41.3.

GENE covy ET PROTEINE COBVSEQUENCE DU FRAGMENT BamHI-SstI-SstI DE 3855 BP, DE 1885 A 2793

109/151

MetLysGlyLeuGlyLysValCysAlaAlaLeuAlaGlyAlaProAlaArgLeuAlaArgLeuAlaAlaValGly  
 ATGAAAGGTCTGGAAAAAGTGTGCGCGGCCCTCGCGGGTGCAGCCGGCTTCGAGACTCGCCGCAGTCGGC  
     1894      1904      1914      1924      1934      1944      1954  
 GluPheGlySerGlyArgValAsnGlyArgAsnLeuGlyGlyAlaMetGlyPheValGlyAspPheCysAspAsp  
 GAGTTTGGCAGCGCCGGGTCAATGGCCGCAACCTTGAGGAGGCCATGGCTTGGCGATTCTGTGACGAT  
     1969      1979      1989      1999      2009      2019      2029  
 ValAlaArgSerIleGlyPheLeuSerArgIleProMetProAlaArgHisPheGluGlyTyrAspGlyArgLeu  
 GTGGCGCGGTGATCGGCTTCAGCCGATCCCGATGCCGTGCGCGTCATTGAGGTATGACGGCGCTTG  
     2044      2054      2064      2074      2084      2094      2104  
 SerArgAlaValArgAlaPheProPheAlaGlyLeuAlaIleAlaLeuProSerAlaAlaValAlaMetAlaLeu  
 AGCCGTGCCGTGCGCGCTTCCCGTTGCCGGCTGGCGATCGCACTGCCATGGCGGCCGTGCCATGGCCCTC  
     2119      2129      2139      2149      2159      2169      2179  
 MetAlaLeuGlnValSerLeuPheAlaAlaPheValValAlaIleGlnAlaLeuValThrGlyAlaLeu  
 ATGGCGCTGCAGGTGAGCTCGCTCTTGCCGCCTTGTGTCGTCGCCATCCAGGCGCTTGTGACGGCGACTG  
     2194      2204      2214      2224      2234      2244      2254  
 HisGluAspGlyLeuGlyAspThrAlaAspGlyPheGlyGlyArgAspArgGluAlaAlaLeuAlaIleMet  
 CACGAAGACGGGCTTGGCGACACGGCCGATGGCTTGCGCGGGTGCACCCGAAAGCGCGCTTGGCATCATG  
     2269      2279      2289      2299      2309      2319      2329  
 LysAspSerArgIleGlyThrTyrAlaAlaValAlaLeuIleLeuSerPheGlyLeuArgValSerAlaPheAla  
 AAGGACAGCCGCATCGGCACCTATGCCGCGCTGCCCTCATCCTCTGCCCTGCCGTCTGGCGTTCGCC  
     2344      2354      2364      2374      2384      2394      2404  
 SerIleLeuProLeuPheSerProLeuGlyAlaAlaMetAlaIleLeuGlyAlaAlaCysLeuSerArgAlaAla  
 TCGATCCTGCCGCTCTTCGCCGCTGGCGCCATGGCGATCCTGCCGAGCGTGCCTCAGCCGCCGCC  
     2419      2429      2439      2449      2459      2469      2479  
 MetValTrpHisTrpSerSerLeuProProAlaArgSerSerGlyValAlaAlaSerAlaGlyGluProGluPro  
 ATGGCTGGCACTGGTCGCTCCGCCGGCGCAGCAGCGCGCTCGCGGCCCTGGCTGGCGAGCCGGAACCG  
     2494      2504      2514      2524      2534      2544      2554  
 AlaAlaThrArgPheAlaLeuAlaPheGlyLeuLeuValAlaMetLeuLeuPheTyrLeuAlaGlnValProAla  
 GCGGCCACCGCTTCGCGCTGCCCTCGGACTTCTGCGCAATGCTGCTCTATCTCGCGCAGGTCCC GGCG  
     2569      2579      2589      2599      2609      2619      2629  
 LeuGlyValIleAlaAlaLeuValAlaPheLeuAlaThrValLysGlyPheAlaArgLeuAlaMetArgLysIle  
 CTCGGCGTCATCGCGCTCGCATTGCCACGGTCAAGGGCTTCGCTGGCGATGCCAAGATC  
     2644      2654      2664      2674      2684      2694      2704  
 GlyGlyGlnThrGlyAspThrIleGlyAlaThrGlnGlnLeuThrGluIleAlaValLeuGlyAlaLeuAlaLeu  
 GGCAGGACAAACGGCGACAGCATGGGGCGACGCAGCAACTGACCGAAATGCCGTGCTCGGTGCCCTGCGCTG  
     2719      2729      2739      2749      2759      2769      2779  
 ThrVal\*\*\*  
 ACGGTTTGA

FIGURE 41.4

PL. 110 / 151

(A)

97.4

66.2

42.7

31



(B)

1 2 3

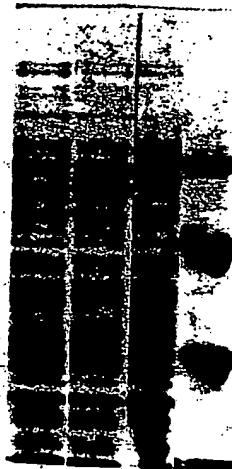


FIGURE 42

FEUILLE DE REMplacement

PL. 111/151

Sequence      Longueur = 13144      de      1 a 13144

10	20	30	40	50	60
GAGCTCGAAG	GGGCTTCCGC	CCCGATCGCT	GGCGTTAGCC	GACGTTGAC	GTGCGGATGA
CTCGAGCTTC	CCCCAAGGCC	GGGCTAGCGA	CCGCAATCGG	CTGCAAGCTG	CACGCCACT
70	80	90	100	110	120
CGCCGAGCGG	GCCGAAGGCC	CGCTCGACGA	CGAGGTTGCG	TACGCGCGAC	TGGCTGGACG
GCGGCTCGCC	GGGCTTCCCG	GGCAGCTGCT	GCTCCAACGC	ATGCGCGCTG	ACCGACCTGC
130	140	150	160	170	180
GAACCTTCGA	GTTCAGGCCG	ATCTGAACGA	AATTGGGCTT	GCTGAAAATA	TACAGCATGG
CTTGGAAAGCT	CAAGGTCCGC	TAGACTTGCT	TTAACCCGAA	CGACTTTAT	ATGTCGTACC
190	200	210	220	230	240
ACATGAACCT	TGAGGAGGCCG	GAGGCCATATC	CTCCGGGCG	TGTTGCTATG	CCGCTGATAT
TGTACTTGGG	ACTCTCCGGC	CTCCGGATAG	GAGGCCCCGC	ACAACGATAC	GGCGACTATA
250	260	270	280	290	300
AGGTGTGCGC	TGCAAAAAAT	TGAATGCCAA	ACTCGCCACG	CCATGTCGCA	TTCTGGCTAT
TCCACACGCG	ACGTTTTTA	ACTTACGGTT	TGAGCGGTG	GGTACAGCGT	AAGACCGATA
310	320	330	340	350	360
CGGCCGCGAC	ATTTTCGACA	AGCCTTGCAG	AAGCGCAA	CAATGCGTGA	AAGGGCTTG
GCCGGCGCTG	AAAAAGCTGT	TGGGAACGCT	TTCCGGCTT	GTTACGCACT	TTCCCGAAAC
370	380	390	400	410	420
TCAATTGCGG	CGAAATCGTG	TCGAAACAGA	CCTTTGCCG	TGCCCCTTTC	AGTGTTACCG
AGTTAACGCC	GCTTTAGCAC	AGCTTTGTCT	GGAAACGGCG	ACGGGCAAAG	TCACAATGGC
430	440	450	460	470	480
ATGGCCGCAT	GACACCGCAGG	ATCATGTTGC	AGGGAAACGG	CTCGGATGTC	GGAAAATCGG
TACCGCGTA	CTGTGCGTCC	TAGTACAACG	TCCCTGGCC	GAGCCTACAG	CCTTTAGCC
490	500	510	520	530	540
TATTGGTGGC	GGGGCTCTGC	CGGCTTGCCT	CCAATCAGGG	CCTGAAGGTC	CGGCCGTTC
ATAACCACCG	CCCCGAGACG	GGCGAACGGC	GGTTAGTCCC	GGACTTCCAG	GCCGGCAAGT
550	560	570	580	590	600
AGCCGCAGAA	CATGTCGAAC	AACGCCGCCG	TTTCCGACGA	CGGCGGCGAG	ATCGGCGCG
TCGGCGCTT	GTACAGCTTG	TTGGCGCGGC	AAAGGCTGCT	GCCGCCGCTC	TAGCCGGCGC
610	620	630	640	650	660
CGCAATGGCT	GCAGGGCGTG	GGCGCGCG	TGCGCTCGTC	GGTGACATG	AACCCGGTGC
GCGTTACCGA	CGTCCGCGAC	CGGCGCGCG	ACGGCAGCAG	CCACGTGTAC	TTGGGCCACCG
670	680	690	700	710	720
TCCCTGAAGCC	GCAGTCGGAC	GTGGGCAGCC	AGATCGCTG	TCAGGGCAAG	GTGCGCCGGC
AGGACTTCGG	CGTCAGCCCTG	CACCCGTCGG	TCTAGCAGCA	AGTCCCCTTC	CAGCGGCCCG
730	740	750	760	770	780
AGGCCAGGGG	GCGGAAATAT	CAGGGCGCTCA	AGCCCAAGCT	GCTGGGCGCC	GTCAATGGAGA
TCCGGTCCCC	CGCCCTTATA	GTCCCGAGT	TCGGGTTCGA	CGACCCGCGG	CAGTACCTCT
790	800	810	820	830	840
GTTTCAAAACA	AAATATCGGCC	GGTCCCGATC	TCGTGGTGGT	CGAAGGGCGCC	GGCTCGCCGG
CAAAGCTTGT	TTATAGCCGG	CCACGGCTAG	AGCACCACCA	GCTCCGCGG	CCGAGCGGCC
850	860	870	880	890	900
CCGAAATCAA	CCTCAGGCC	GGCGACATCG	CCAATATGGG	CITTCGACAA	CGGGCCAATG
GGCTTTAGTT	GGAGTCCGGG	CCCGCTGTAGC	GGTTATACCC	GAAACGCTGT	CCCCGGTTAC
910	920	930	940	950	960

FIGURE 43

PL. 112 / 151

TGCCGGTCGT GCTGGTCGGC GACATCGACC GCGGGGGGGT GATCGCCTCG CTGGTCGGCA  
 ACGGCCAGCA CGACCAGCG CTGTAGCTGG CGCCCCCCC CTAGCGGAGC GACCAGCCGT  
 970 980 990 1000 1010 1020  
 CGCATGCGAT CCTGCCCAGA GAAGACCGGC GCATGGTGAC CGGCTATCTC ATCAACAAGT  
 GCGTACGCTA GGACGGGCTC CTTCTGGCCG CGTACCAACTG GCGGATAGAG TAGTTGTTCA  
 1030 1040 1050 1060 1070 1080  
 TCCCGGGCGA CGTCACCGCTG TTCGACGACG GCATTGCTGC CGTCAACCGC TACACCGCT  
 AGGCGCCGCT GCAGTGCAGC AAGCTGCTGC CGTAACCGACG GCAGTTGGCG ATGTTGGCCGA  
 1090 1100 1110 1120 1130 1140  
 GGCCCTGCTT CGGCCTCGTG CGCGGGCTGA AGGCGGGCGC ACGGCTGCGC GCGGAAGATT  
 CCGGGACGAA GCCGACGAC GGCACCGACT TCCGCCCGC TGCGGACGGC CGCCTTCTAA  
 1150 1160 1170 1180 1190 1200  
 CCGTCGTGCT GGAGAACGCTG ACGCGCGGGC AGGGCGGGC GCTGAAGGTT GCCGTCGGGG  
 GGCAGCACGA CCTCTTCGAC TGCGCGCCCG TCCCCCGCCG CGACTTCCAA CGGCAGGGCC  
 1210 1220 1230 1240 1250 1260  
 TACTGTCGCG CATCGCAAAT TTGACGACCC TCGATCCGCT CGCCGCGGAA CGGGAGATTG  
 ATGACAGCGC GTAGCGGTTA AAGCTGCTGG AGCTAGGCGA GCGGCGGCTT GGCCTCTAAC  
 1270 1280 1290 1300 1310 1320  
 ATCTCGTCTT CGTGCAGGCTT GGCAAGTCCC TTCCGGTCGA CGCTGGCCTC GTCGTCATTC  
 TAGAGCAGAA GCACGCCGA CGTCAGGGT AAGGCCAGCT GCGACCGGAG CAGCAGTAAG  
 1330 1340 1350 1360 1370 1380  
 CGGGGTCGAA ATCGACCATC GGCGACCTCA TCGATTCCG TGCGCAAGGG TGGGACCGTG  
 GGCCCAGCTT TAGCTGGTAG CGCGTGGAGT AGCTAAAGGC ACGCCTTCCC ACCCTGGCAC  
 1390 1400 1410 1420 1430 1440  
 ACCTCGAACG TCATGTCGCG CGGGGCGGGC GGGTCATCGG CATCTGCGGC GGCTACCAAGA  
 TGGAGCTTGC AGTACACCGC GCCCCGCCGG CCCAGTAGCC GTAGACGCCG CCGATGGTCT  
 1450 1460 1470 1480 1490 1500  
 TGCTCGGCCG GCGCGTCACC GATCCGCTCG GCATCGAGGG CGGGGAACGT GCGGTCGAGG  
 ACGAGCCGGC CGCGCAGTGG CTAGGCGAGC CGTAGCTCCC GCGCCTTGCA CGCCAGCTCC  
 1510 1520 1530 1540 1550 1560  
 GCCTCGGGCT GCTCGAGGTC GAGACCGAGA TGGCGCCGGA AAAGACGGTG CGCAACAGCC  
 CGGAGCCCGA CGAGCTCCAG CTCTGGCTCT ACCCGGGCCT TTCTGCCAC GCGTTGTCGG  
 1570 1580 1590 1600 1610 1620  
 GCGCCTGGTC GCTGGAGCAT GATGTGGTGC TCGAAGGCTA CGAAATCCAT CTTGGCAAGA  
 CGCGGACCAAG CGACCTCGTA CTACACCACG AGCTTCCGAT GCTTAAAGGTA GAACCGTTCT  
 1630 1640 1650 1660 1670 1680  
 CGCAAGGTGC GGACTGTCGGC CGGGCGTCGG TCGCGCATCGA CAATCGCGCC GACGGCGCCC  
 CGCTTCCACG CCTGACACCG CGCGGCAGCC AGCGTIAAGCT GTAGCGCGGG CTGCCGCCGG  
 1690 1700 1710 1720 1730 1740  
 TTTCGGCCGA TGCGCGCTG ATGGGCACCT ACCTGCATGG GCTCTTCACC AGCGACGCC  
 AAAGCCGGCT ACCGGCGCAC TACCCGTGGA TGGACGTACC CGAGAAGTGG TCGCTGCGGA  
 1750 1760 1770 1780 1790 1800  
 ATCGCGGCCGC GCTGCTCAAG AGTTTCGGCA TCGAAGGGCGG CGCCAACAAAC TACCGCCAAT  
 TAGCGCCCGCG CGACGAGITC TCAAAGCCGT AGCTTCCGCC GCGGTTGTTG ATGGCGGTTA  
 1810 1820 1830 1840 1850 1860  
 CGGTGATGC GCGCGTCGAC GATGTGCGCA ACCAACCTGGA GGCTGTGCTC GATCGTCGCT  
 GCCAGCTACG CGCGCAGCTG CTACAGCGCT TGCTTGACCT CGACACCGAG CTAGCAGCGA  
 1870 1880 1890 1900 1910 1920  
 GGCTGGACGA GTTGCTCAGG CACTAGGGAC GCGGCAACGG TCAGCCAGCA GGTCCGGTAC

FIGURE 43.1

PL. 113 / 15 /

CCGACCTGCT CAACGAGTCC GTGATCCCTG CGCCGTGCG AGTCGGTCGT CCAGGCCATG  
 1930 1940 1950 1960 1970 1980  
 GTCGGGCCCA ACAGGAGCAA CGAGCTTATC CGACGGAACT ACCGCTGCCAC ATCGTGCTCC  
 CAGCCCCGGT TGTCTCGIT GCTCGAATAG GCTGCCTTGA TCGCACGCTG TAGCACGAGG  
 1990 2000 2010 2020 2030 2040  
 TCGCTTGGCG CTTCCCAGAC TTCCCGCCCG GCATCCAGGT TCATCAGGGC AATCCCCAGG  
 AGCGAACGCC GAAGGGTCTG AAGGGCGCGC CGTAGGTCCA AGTAGTCCCG TTAGGGGTCC  
 2050 2060 2070 2080 2090 2100  
 CCGACCATCA GGTCGGGCCA GGCGACTGC CACAGATAGG CTGTCGCCAG ACCCGCGGGCG  
 GGCTGCTAGT CCAGGCCGGT CCGGTGACG GTGCTATCC GACAGCGTC TGGCGCCCGC  
 2110 2120 2130 2140 2150 2160  
 ATGATGGCCA CATTGGCGAA GGCATCGTTG CGGGCGGAGA GAAATGCTGC CCGCGTGAGC  
 TACTACCGGT GTAAACCGCTT CCGTAGCAAC GCGCGCTCT CTTACGACG GCGCACTCG  
 2170 2180 2190 2200 2210 2220  
 GTGCCGCTCG TGTGACGGTA GGCGACGAGC AGATAGGCAG AGAAGAGGTT GACCACCAGC  
 CACGGCGAGC ACACTGCCAT CCGCTGCTG TCTATCCCGC TCTTCTCCAA CTGGTGGTCC  
 2230 2240 2250 2260 2270 2280  
 GCCCCAAGTC CGGTCAAGGGA CAGGGCAAGG GGCTCTGGCG GGACCGGATC CATGAACCTTC  
 CGGGGTTCAAGC GCCAGTCCCT GTCCCGTTTC CCGAGACCGC CCTGGCCTAG GTACTTGAAG  
 2290 2300 2310 2320 2330 2340  
 GCCCAGGCGC TCCAAAGGAA GGCCAGCGC GGTACCAAGTA GAATGAACGC CATGCCATG  
 CGGGTCCGGC AGGTTTCCCTT CCGGTGCGG CCATGGTCAT CTACTTGCG GTAGCGGTAC  
 2350 2360 2370 2380 2390 2400  
 CCGACCCGCG CGCGGGTTCA CGCCGTCCAG GCGAGAGCAA AGAAAATCAG CATGTTGACG  
 GGCTGGCGC GCGCCCAAGC GCGGCAAGTC CGGTCTCGTT TCCTTCTAGTC GTACAACCTGC  
 2410 2420 2430 2440 2450 2460  
 GAGGCGTCCT CGAGGAAGTC GACGTGTCG GCCATGAGGG ACACCGAGCC GATCGAAAGC  
 CTCCGCAAGA GCTCTTCAG CTGCGACAGC CGGTACTCCC TGTTGGCTCGG CTAGCTTTCG  
 2470 2480 2490 2500 2510 2520  
 GCGACAAGGA GTTCGACCCC GAAATAGCCA AGGTTCAACA GGGAGACGAT GAGGACGACG  
 CGCTGTTCTT CAAGCTGGGG CTTTATCGGT TCCAAGTTGT CCCCTCTGCTA CTCTGCTGC  
 2530 2540 2550 2560 2570 2580  
 CGGCGCAGGT CGGTATCCAC TCGAAAGGTT CCCTTTCTGG CGAGATTGCG CTCGGCACT  
 GCGCGTCCA GCCATAGGTG AGCTTCCAA GGGAAAGACC GCTCTAAGCG GGAGCCGTGA  
 2590 2600 2610 2620 2630 2640  
 TTTTTGGCG AGATTCGCC CGGGCACTTT GGCACAGGTG TTAGCAGCAG TTTGCTATCC  
 AAAAAACCGC TCTAAGCGGG AGCCGTGAAA CCGTGTCCAC AATCGTCGTC AAACGATAGG  
 2650 2660 2670 2680 2690 2700  
 ATAGCACTAG GTTTGACAT CGGTTCCCGT CACACTGCCG TCGTGCCTGA CGCCCGACAA  
 TATCGTGATC CAAAGCTGTA GCCAAGGCAA GTGTGACGGC AGCACGGACT CGGGGCTGTT  
 2710 2720 2730 2740 2750 2760  
 ATCGTCGCGT GGCACAACTC GGCGGGGGAG GCGTCGCATG CGTCGATTGA CTTTGGGCTG  
 TAGCAGCGCA CGCGGTGAG CGGGCCCCCTC CGCAGCGTAC GCAGCTAAGT GAAACCCGAC  
 2770 2780 2790 2800 2810 2820  
 CCCGCTTCCTT AATCATCAGG TGTGGATGG TTCCCGCTTG TCGTGGCGAT CTGGGGGAAT  
 GGGCGAAGGA TTAGTAGTCC ACAAACCTACC AGGGGGAAAC AGCACCGCTA GACCCCTTA  
 2830 2840 2850 2860 2870 2880  
 AATTGGGAAT GTGACGGATG GACCCAAATC GGGCTCCCTT ATCGCAAGCCG ACCCCGGAC  
 TTAACCCCTTA CACTGCTAC CTGGGTTAG CCCGTAGGAA TAGCGTCGGC TGGGGCGCTG

FIGURE 43.2

FEUILLE DE REMplacement

PL. 114 [15]

2890	2900	2910	2920	2930	2940
TGTAGAACGG	TCAGGGITCG	CCATCGGGAT	TGGTGCAGGG	CTGTCGGCCG	GTTGCATGGG
ACATCTTGCC	AGTCCCAAGC	GGTAGGCCCTA	ACCACGGCCC	GACAGCCGGC	CAACGTACCC
2950	2960	2970	2980	2990	3000
CAATCGGGGC	AGGTGGGGA	TCAAGCCGGA	AAAGCCACTG	GCGTGGCATC	GTGATCAGCC
GTTAGCCCCG	TCCAGCCCC	AGTTCGGCCT	TTTCGGTGAC	CGCACCGTAG	CACTAGTCGG
3010	3020	3030	3040	3050	3060
GGGTTGGAC	GCCTCTTCCT	CTACGAATCG	TCCGCCTTTC	ACGATGTCCC	TCACAGCGCC
CCCAAACCTG	CGGAGAAGAA	GATGCTTAGE	AGGCGGAAAG	TGCTACAGGG	AGTGTGCGGG
3070	3080	3090	3100	3110	3120
CATCGGTCGG	AGACGACGCG	CAAAGGTTCG	CTGTCGGCACC	GGAAAGACGC	CGGGAAAGGTG
GTACGCAGCC	TCTGTCGGC	GTTTCCAAGC	GACACCGTGG	CCTTTCTGCG	GCCCTTCCAC
3130	3140	3150	3160	3170	3180
AGGCGGGCCG	CTCGGGCCCT	GACATCGGAA	CCTTGCCTGT	TAAGGGCGAG	GCGATGTTCG
TCCGCCGGC	GAGCCCGGG	CTGTAGCCTT	GGAACGGCAA	ATTCCCGCTC	CGCTACAAGC
3190	3200	3210	3220	3230	3240
GCCCCGTGACG	CCGTGAGCCA	GGAGACCTGC	CATCCGGCAT	GGGCATTCGG	CCCCAGGGGA
CGGGCACTGC	GGCACTCGGT	CCTCTGGACG	GTAGGCCGTA	CCCGTAAGGC	GGGCTCCCC
3250	3260	3270	3280	3290	3300
CTTTTGTCCTC	CAACGCCATC	ACGGAGGTG	TTTGGCTCG	CAGATGTTTT	CAAGAACGCG
AAAAACAGAG	GTTGCGTAG	TGCCCTCCAC	AAAACCGAGC	GTCACAAAA	TTTCTTGCAC
3310	3320	3330	3340	3350	3360
CCCGTGGCGC	GTCCGATGGC	TTTGGCCACC	GACGGCTGAT	TTGGGAATGT	TGAGGGCAGCC
GGGCACCGCG	CAGGCTACCG	AAAACGGTGG	CTGCCGACTA	AACCCTTACA	ACTCCGTCGG
3370	3380	3390	3400	3410	3420
ACGATGAGCA	GTCTCAGCGC	CGGGCCCCGTG	CTGGTCCTTG	GCAGCGCCCG	TTCCGGCAAG
TGCTACTCGT	CAGAGTCGCG	GCCCCGGCAC	GACCAGGAAC	CGCCCGGGC	AAGGCCGTT
3430	3440	3450	3460	3470	3480
TCCAGCTTT	CCGAGAGGCT	CCTCGAAGCG	TCCGGCTTCA	CCATGCATTA	TGTCGCCACG
AGGTCGAAA	GGCTCTCCGA	GCAGCTTCGC	AGGCCGAAGT	GGTACGTAAT	ACAGCGGTGC
3490	3500	3510	3520	3530	3540
GGCCGCCGCT	GGGACGACGA	AATGCGCGAG	CGCATCGACC	ATCACCGGAC	GCGCCGCGGC
CCGGCGCGGA	CCCTGCTGCT	TTACCGCGTC	GCGTAGCTGG	TAGTGGCCTG	CGCGGGCGCCG
3550	3560	3570	3580	3590	3600
GAGGGCTGGA	CGACGCATGA	GGAGCCGCTC	GATCTCGTCG	GCATCCTCAG	ACGCATCGAT
CTCCCGACCT	GCTGCGTACT	CCTCGCGAG	CTAGAGCAGC	CGTAGGAGTC	TGCGTAGCTA
3610	3620	3630	3640	3650	3660
GATCCCAAGCC	ATGTTGGCTC	GATCGACTGC	CCTGACGCTAT	GGGTACCAA	TCTCATGCTG
CTAGGGTCGG	TACACCAGGA	CTAGCTGACG	GACTCGCATA	CCCACTGGTT	AGAGTACGAC
3670	3680	3690	3700	3710	3720
GAAGAGCGCG	ACATGACGGC	GGAGTTGCGC	GCCCTTGTG	CGTACTGCCC	CGAGGGCGCG
CTTCTCGCG	TGTACTGCCG	CCTCAAGCGG	CGGGAAACAAC	GCATAGACGG	GCTCCGCGCC
3730	3740	3750	3760	3770	3780
GCGCGCCCTCG	TCTTTGTTTC	CAATGAGGTC	GGCCTCGGCA	TCGTGCCCCG	GAACCGCATG
CGCGCGGGAGC	AGAACAAAG	GTTACTCCAG	CGGGAGCGT	AGCAACGGCT	CTTGGCGTAC
3790	3800	3810	3820	3830	3840
GCCCCGCGAGT	TTCGCGACCA	TGCGGGCCGG	CTTCACCAAGA	TCGTGCGGA	GAATCCGCT
CGGGCGCTCA	AAGCGCTGGT	ACGGCCGGCC	GAAGTGGTCT	ACCAACGCC	CTTTAGGCAG

FIGURE 43.3

PL. 115 / 151

3850	3860	3870	3880	3890	3900
GAAGTTACT TTGCGCGGC CGGTTGCCG CTGAAAATGA AGGGTTGATC CATGACCACT CTTCAAATGA AACAGCGCCG CCAAACGGC GACTTTACT TCCCAACTAG GTACTGGTGA					
3910	3920	3930	3940	3950	3960
GCGAGAGCCA ACCAGGGCAA GATCCCCGGCG ACCGTCACTA CCGGCTTCCT CGGCGCCGGC CGCTCTCGGT TGGTCCCCTT CTAGGGCCGC TGCGAGTAGT GGCGAAGGA GCCGCGGGCCG					
3970	3980	3990	4000	4010	4020
AAGACGACGA TGATCCGCAA CCTGCTGCAG AACGCCGACG GCAAGCGCAT CGGCCTGATC TTCTGCTGCT ACTAGGCGTT GGAGGACGTC TTGCGGCIGC CGTTCGCGTA GCCGGACTAG					
4030	4040	4050	4060	4070	4080
ATCAAACGAGT TCGGCGATCT TGGCGTCGAC GGCGATGTC TGAAAGGGCTG CGGTGCGAG TAGTTGCTCA AGCCGCTAGA ACCGGCAGCTG CGCGTACAGA ACTTCCCGAC GCCACGGCTC					
4090	4100	4110	4120	4130	4140
GCCTGCACCG AGGACGACAT CATCGAGCTC ACCAATGGCT GCATCTGCTG CACCGTGGCT CGGACGTGGC TCCTGCTGTA GTAGCTCGAG TGGTACCGA CGTAGACGAC GTGGCACCGA					
4150	4160	4170	4180	4190	4200
GACGATTTC A TCCCGACCAT GACGAAGCTG CTCGAGCGTG AAAACCGTCC TGACCACATC CTGCTAAAGT AGGGCTGGTA CTGCTTCGAC GAGCTCGCAC TTTTGGCAGG ACTGGTGTAG					
4210	4220	4230	4240	4250	4260
ATCATCGAAA CCTCGGGCCT TGCCCTGCCG CAGCCGCTGA TCGCCGCTTT CAACTGGCCG TAGTAGCTTT GGAGCCCGGA ACAGGACGGC GTCGGGACT AGCGCGAAA GTTGACCGGC					
4270	4280	4290	4300	4310	4320
GATATCCGCA GCGAAGTGAC CGTCGATGGC GTCGTCACCG TGGTCGACAG CGCCGCGTT CTATAGGCGT CGCTTCACTG GCAGCTACCG CAGCAGTGGC ACCAGCTGTC GCGGCGGCAA					
4330	4340	4350	4360	4370	4380
GCCGCTGGCC GCTTTGCCGA CGACCACGAC AAGGTGCGATG CGCTGCGGT CGAGGACGAC CGGCGACCGG CGAACACGGCT GCTGGTGCTG TTCCAGCTAC GCGACGCGCA GCTCCTGCTG					
4390	4400	4410	4420	4430	4440
AATCTCGATC ACGAAAGCCC GATCGAGGAG CTGTTGAGG ATCAACTGAC GGCTGCCGAT TTAGAGCTAG TGCTTTCGGG CTAGCTCCTC GACAAGCTCC TAGTTGACTG CCGACGGCTA					
4450	4460	4470	4480	4490	4500
CTCATCGTC TCAAACAAGAC CGATCTGATC GATGCCCTCG GCCTCAAGGC CGTGCACGAC GAGTAGCAAG AGTTGTTCTG GCTAGACTAG CTACGGAGGC CGGAGTTCCG GCACGCGCTG					
4510	4520	4530	4540	4550	4560
GAGGTGCTTT CGCGCACCAAG CGCGAAGCCC ACGATGATCG AGGCAAAAAA CGGCGAAGTC CTCCACAGAA GCGCGTGGTC GCGCTTCGGG TGCTACTAGC TCCGCTTTT GCGCTTCAG					
4570	4580	4590	4600	4610	4620
GCCGCTGCA TCCGTCTTGG CCTCGGTGTC GGCACGGAAA GCGATATCGC CAACCGCAAG CGGCGACGGT AGGACGAACC GGAGCCACAG CCGTGCTTT CGCTATAGCG GTTGGCGTTC					
4630	4640	4650	4660	4670	4680
TCGCATCAGC AGATGGAGCA CGAGGCGAGT GAGGAGCACG ATCACGACGA GTTCGACAGC AGCGTAGTGC TCTACCTCGT GCTCCGTCCA CCTCTCGTGC TAGTGTGCT CAAGCTGTGC					
4690	4700	4710	4720	4730	4740
TTCGTCGTGAGCTCGGTT GATCGCCGAT CGGGCCGCCT TCATCGATCG CCTGAAGGGC AAGCAGCAGC TCGAGCCAAG CTAGCGCTA GGCGGGCGGA AGTAGCTAGC GGACTTCCCG					
4750	4760	4770	4780	4790	4800
GTAATCGCGG AGCACGACGT TCTGCGCTTC AAGGGTTTTG CAGACGTGCC CGGCAAGCCG CAATTAGCGCC TCGTGTGCA AGACGCGGAG TTCCCAAAAC GTCTGCACGG GCGGTTCGGC					
4810	4820	4830	4840	4850	4860

FIGURE 43.4

(PL. 116 / 15)

ATGCGCCCTCC TGATCCAGGC GGTGGCGGCC CGCATCGACC AATAATTACGA CCGCGCCTGG  
 TACGCGGAGG ACTAGGTCCG CCAGCGCGG GCGTAGCTGG TTATAATGCT GGCAGCGGAC  
 4870 4880 4890 4900 4910 4920  
 GGGCGTGGCG AAAAGCCGG TACCGGCCCTC GTCGTCATCG GCCTGCACGA CATGGACGAG  
 CCGCGACCGC TTTTCGCGCC ATGCGCGGAG CAGCAGTAGC CGGACGTGCT GTACCTGCTC  
 4930 4940 4950 4960 4970 4980  
 GCGGCGGTGCG CCGCCGCGAT CACCGCGCTC GTGTAGATCG TTCTTTGAAT GAAATGATCT  
 CGCCGCCACG CGCGCGCTA GTGGCGCGAG CACATCTAGC AAGAAAATTA CTTTACTAGA  
 4990 5000 5010 5020 5030 5040  
 AACGCATTGA AATGATGCAAG TTCCGGATGG AGAACGCTTT TAGCGTTTC GTTCGGAATT  
 TTGGCTAATC TTACTACGTC AAGGCCCTAC TCTTGCAGAA ATCCAAAAG CAAGCCTTAA  
 5050 5060 5070 5080 5090 5100  
 GCCCCAACCGG ACAAGACGAA TGCATCTGCT TCTCGCCAG AAAGGAACGA TCGCCGACGG  
 CGGGGTTGCC TGTTCGCTT ACCTAGACGA AGAGCGGGTC TTTCCTTGCT AGCGGCTGCG  
 5110 5120 5130 5140 5150 5160  
 CAACGAGGCG ATCGACCTTG GGCAAAACGCC GGCGATATC CTTTTCCAT CGGCGGCCGA  
 GTTGCCTCCGC TAGCTGGAAC CGGTTTGCGG CGGGCTATAG GAAAAGGATA CCCGGCGGCGT  
 5170 5180 5190 5200 5210 5220  
 CACCGAGCTC TCCTCGATCG CGCGGGCTCA CGGCCGACGC GACGGAGGCT TGAGCCTGCG  
 GTGGCTCGAG AGGAGCTAGC GGCGCCGAGT GCCGGCTGCG CTGCCCTCGA ACTCGGACGC  
 5230 5240 5250 5260 5270 5280  
 CATCGCCAGC CTGATGAGCC TGATGCCACCC GATGTGGTC GACACTTACG TCGAGCGCAC  
 GTAGCGGTGCG GACTACTCGG ACTACGTGGG CTACAGCCAG CTGTGAATGC AGCTCGCGTG  
 5290 5300 5310 5320 5330 5340  
 GGCGCGTCAC GCCAAGCTGA TCGTCGTCGG GCGCTCGGT GGCGCCAGCT ATTTCGTTA  
 CCGCGCAGTG CGGTTGACT AGCAGCAGGC CGCGAGCCA CGCGGCTCGA TAAAGGCAAT  
 5350 5360 5370 5380 5390 5400  
 TCTGCTGGAA GCCCTGCATG CGGCTGCCGT CACCCATCGT TTGAGATCG CGGTTCTGCC  
 AGACGACCTT CGGGACGTAC GCGACGGCA GTGGGTAGCA AAGCTCTAGC GCCAAGACGG  
 5410 5420 5430 5440 5450 5460  
 GGGTGACGAC AAGCCGGATC CGGGGCTGGG GCCTTCTCC ACCGTCGCGAG CCGACGACCG  
 CCCACTGCTG TTCCGGCTAG GCCCCGACCT CGGAAAGAGG TGGCAGCGTC GGCTGCTGGC  
 5470 5480 5490 5500 5510 5520  
 CCAGCGCCCTT TGGGCTTACT TCACCGAAGG CGGCTCGGAC AATGCCGGGC TGTTTCTCGA  
 GGTCGCGGAA ACCCGAATGA AGTGGCTTCC CGCGAGCTG TTACGGCCCG ACAAAGAGCT  
 5530 5540 5550 5560 5570 5580  
 CTATCGGGCC GCACTGGTCA CAGGTGGCGGA GAAGCCGCAG CGGCAAAAGC CCCTGTTGAA  
 GATACGCCGG CGTGACCACTG GTCCACGCGCT CTTGGCGTC CGCGGTTTCG GGGACAACCT  
 5590 5600 5610 5620 5630 5640  
 GGCGGGCATC TGGTGGCCGG GTGCTGGTGT GATCGGGCGC AGCGAATGGC AGTCCCTTGT  
 CGGGCCGTAG ACCACCGGGCC CACGACCAACA CTAGCGCGAG TCGCTTACCG TCAGGGAACA  
 5650 5660 5670 5680 5690 5700  
 TCAGGGACGG ATGGTAGCGA GGGAGGGATT CGAACCCCCG ACGGTCGGGA TCTGCTTTA  
 AGTCCCTGCGC TACCATCGCT CCCTCCCTAA GCTTGGGGC TGCCAGCCCT AGACGAAAT  
 5710 5720 5730 5740 5750 5760  
 CGCGCGCGTC GTGCAGAGTG GCGAGACACG GCCTGTTGGAG GCGCTGATCG ATGCGCTGG  
 CGCGCGCGAG CACGTCTCAC CGCTCTGTGC CGGACACCTC CGCGACTAGC TACGCGACCT  
 5770 5780 5790 5800 5810 5820  
 GGCTGAAGGT GTGCGGGCAC TGCCGGTGT TGTCTCAAGC CTCAAGGATG CGGTTCCGT

FIGURE 43.5

PL. 117/151

CCGACTTCCA CACGCCCGTG ACGGCCACAA ACAGAGTCG GAGTTCCCTAC GGCAAAAGGA  
 5830 5840 5850 5860 5870 5880  
 CGGCACGCTG CAGGCATT TTTCCGAGCG CGCACCCGAC GTGGTGTGATGA ACGCCACTGG  
 GCCGTGCGAC GTCCGCTAAA AAAGGCTCCG GCGTGGCTG CACCACTACT TGCGGTGACC  
 5890 5900 5910 5920 5930 5940  
 CTTTGGGTC TCGTCGCCC GTGCCGACCG TCAGCCGACG GTGCTGGAAT CGACCGGTGC  
 GAAACGCCAG AGCAGCGGGC CACGGCTGGC AGTCGGCTGC CACGACCTTA GCTGGCCACG  
 5950 5960 5970 5980 5990 6000  
 CGCGGTGCTG CAGGTGATT TCTCCGGCTC GTCGCGGGCG CAATGGGAAA CGTCGCGGCAC  
 CGGCCACGAC GTCCACTAAA AGAGGCCAG CAGGCCCGC GTTACCCCTT GCAGCGCGGT  
 6010 6020 6030 6040 6050 6060  
 GGGGCTGATG GCGCGCGACCG TCGCCATGAA CGTGGCACTC CCCGGAAGTCG ATGGCCGCAT  
 CCCCCGACTAC CGCGCGCTGG AGCGGTACTT GCACCGTGAG GGGCTTCAGC TACCGCGTA  
 6070 6080 6090 6100 6110 6120  
 CCTTGCACGC GCGCTCTCT TCAAGGCCG GTCGATCTAT GACGCCAAGG TGGAGGCCAA  
 GGAACCGCGC CGGCAGAGGA AGITCCGCCG CAGCTAGATA CTGCGGTTCC ACCTCCGGTT  
 6130 6140 6150 6160 6170 6180  
 TATCGTCGGC CATGAGCCGC TCGAAGGCCG GGTGCGCTT GCGCCTGATC TTGCGGTCAA  
 ATAGCAGCCG GTACTCGGC AGCTTCCGGC CCACGCCAAA CGGGCACTAG AACGGCAGTT  
 6190 6200 6210 6220 6230 6240  
 CTGGGCGAAC GTGCCCGGGG CAGAGCCCCG CGAGCGCCGT ATTGCCATCG TCATGGCCAA  
 GACCCGCTTG CACGCCGGCC GCTCGGGCG GCTCGCCGA TAACGGTAGC AGTACCGGTT  
 6250 6260 6270 6280 6290 6300  
 CTATCCGAAAC CGCGACGGTC GCCTCGGCAA CGGTGTCGGG CTGGACACGC CGGCCGGTAC  
 GATAGGCTTG GCGCTGCCAG CGGAGCCGTT GCCACAGCCC GAGCTGTGCG GCCGGCCATG  
 6310 6320 6330 6340 6350 6360  
 CGTCGAGGTG CTTAGCGCCA TGGCGCGGA AGGCTATGCG GTCGGTGAGG TTCCCGCCGA  
 GCAGCTCCAC GAATCGGGT ACCGCGCCCT TCCGATACGC CAGCCACTCC AAGGGCGGGCT  
 6370 6380 6390 6400 6410 6420  
 TGGCGACGCG CTGATCCGCT TTCTGATGGC CGGGCCGACC AATGCGGCCGA GCCATGACCG  
 ACCGCTGCGC GACTAGGCCA AAGACTACCG GCCCCGGCTGG TTACGCGCT CGGTACTGGC  
 6430 6440 6450 6460 6470 6480  
 TGAAATCCGC GAGCGTATTT CGCTGAACGA TTACAAAACG TTCTTCGATT CGCTTCCGAA  
 ACTTTAGGCG CTCGCATAAA GCGACTTGCT AATGTTTGC AAGAAGCTAA CGGAAGGCTT  
 6490 6500 6510 6520 6530 6540  
 ACAGATAAAAG GATGAAGTGT CGGGTCGCTG GGGCGTGCAG GAGGCCGATC CCTTTTCCT  
 TGTCTATTTC CTACTTCAAC GGCCAGCGAC CCCGCACGGC CTCCGGCTAG GAAAAAGGAA  
 6550 6560 6570 6580 6590 6600  
 CGATGGCGCC TTCGCGCTGC CGCTCGCCCG CTTCGGCGAG GTGATCGTCG GCATCCAACC  
 GCTACCGCGG AAGCGCGACG GCGAGCGGGC GAAGCCGCTC CACTAGCAGC CGTAGGTTGG  
 6610 6620 6630 6640 6650 6660  
 GGCAGCGCCGC TACAACATCG ATCCGAAGGA AAGCTACCAT TCCCGGGACC TCGTGCGGCC  
 CGCGCGCGCG ATGTTGTAGC TAGGCTTCC TTCGATGGTA AGGGGCTGG AGCACGGCG  
 6670 6680 6690 6700 6710 6720  
 GCATGGCTAT CTGGCTTCT ACGCTTCCT CGGCCAGCAG TTGGAGCGC AGGCCGATCGT  
 CGTACCGATA GAGCGGAAGA TGCGGAAGGA CGCGGTGTC AAGCCTCGCG TCCGCTAGCA  
 6730 6740 6750 6760 6770 6780  
 CCACATGGGC AAGCACGGCA ATCTCGAATG GCTGCCGGGC AAGGCCGCTGG CGCTGTGGA  
 GGTGTACCGG TTCGTCGCGT TAGAGCTTAC CGACGGCCCG TTCCGCGACCG CCGACAGCCT

FIGURE 43.6

FEUILLE DE REMPLACEMENT

PL. 118 / 151

6790	6800	6810	6820	6830	6840
AACCTGCTAT	CCCGAACGCA	TCTTCGGGCC	GCTGCCGCAC	ATCTATCCCT	TCATCGTCAA
TTGGACGATA	GGGCTTCGCT	AGAAGCCCCG	CGACGGCGTG	TAGATAGGGA	AGTAGCAGTT
6850	6860	6870	6880	6890	6900
CGATCCGGCG	GAAGGTACGC	AGGCCAACGCG	CCGCACCAGC	GCCGTCATCA	TCGACCACCT
GCTAGGCCG	CTTCCATCGC	TCCGGTTCGC	GGCGTGGTCG	CGGCAGTAGT	AGCTGGTGGA
6910	6920	6930	6940	6950	6960
GACCCCGCCC	TTGACCGCGC	CCGAATCTTA	CGGCCCCGCTC	AAGGATCTGG	AAGCGCTCGT
CTGGGGCGGG	AACTGCGCGC	GGCTTAGGAT	GCCGGGGCGAG	TTCCTAGACC	ITCGCGAGCA
6970	6980	6990	7000	7010	7020
CGACGAATAT	TACGACGCCG	CCGGCGGTGA	TCCGCGCCGC	CTCAGGCTGC	TCAGCCGCCA
GCTGCTTATA	ATGCTGCGGC	GGCCGCCACT	AGGCGCGGCG	GAGTCCGACG	AGTCGGCGGT
7030	7040	7050	7060	7070	7080
GATCCCTCGAT	CTCGTGC CGCG	ACATCGGCCT	CGACAGCGAC	GCAGGCATCG	ACAGGGGCGA
CTAGGAGCTA	GAGCACCGCGC	TGTAGCCGGA	GCTGTCGCTG	CGTCCTGTAGC	TGTCCCCGCT
7090	7100	7110	7120	7130	7140
CAGCGACGAC	AAGGCGCTGG	AAAAGCTCGA	CGCCTATCTC	TGCGACCTCA	AGGAAATGCA
GTCGCTGCTG	TTCCCGACCC	TTTTCGAGCT	GCGGATAGAG	ACGCTGGAGT	TCCTTACGT
7150	7160	7170	7180	7190	7200
GATCCCGCGAC	GGCCCTGCACA	TCTTCGGCGT	TGCGCCGGAA	GGGGCGGTGT	TGACGGACCT
CTAGGCGCTG	CCGGACGTGT	AGAAGCCGCA	ACGCGCCCTT	CCCGCCAACA	ACTGCCCTGGA
7210	7220	7230	7240	7250	7260
CACCGTAGCG	CTGGCGCGCG	TGCCCCGAGG	TCTCGGGCAG	GGCGCGGACCC	AGAGCCTGCA
GTGGCATCGC	GACCGCGCGC	ACGGGGCTCC	AGAGCGCTC	CCGGCGCTGG	TCTCGGACGT
7270	7280	7290	7300	7310	7320
CGGGCGATC	GCAGCGGATG	CGGGGCTGCG	TGGGTTTGCT	ATTCCACCT	CGGCGGGGGG
CGCCCCCTAG	CGTCGCTTAC	CCCCCGACCG	ACCCAAACGA	TAAGGGTGGA	CCCGCCCCCC
7330	7340	7350	7360	7370	7380
CAACCCCGCA	CGCGACGCC	AACCTTCGA	CCCGCTCGAC	TGCGTCATGT	CCGACACCTG
GTTGGGGCGT	GCGCTGC GGG	TTGGGAAGCT	GGGCAGCTG	ACGCA GTACA	GGCTGTGGAC
7390	7400	7410	7420	7430	7440
GACAGGCCCG	AAACCGTCCA	TCCTCGCTGA	CCTCTCGGAC	GCCCCCTGGC	GCACCGCCGG
CTGTCCGGGC	TTTGGCAGGT	AGGAGCGACT	GGAGAGCCTG	GGGGGGACCG	CGTGGCGGGCC
7450	7460	7470	7480	7490	7500
CGATACGGTC	GAGCGCATCG	AGTTGCTTGC	CGAAATCTC	GTGTCGGGTG	AACTGGCTTG
GCTATGCCAG	CTCGCGTAGC	TCAACGAACG	GGCTTGTAGAG	CACAGCCCAC	TTGACCGAAC
7510	7520	7530	7540	7550	7560
CCCGGACAC	TGGGCCAACA	CCCGCGCCGT	GCTCGCGAA	ATCGAAACGC	GCCTGAAGCC
GGGCCTGGTG	ACCCGGTTGT	GGGC GCGGCA	CGAGCGCCTT	TAGCTTTGCG	CGGACTTCGG
7570	7580	7590	7600	7610	7620
GTCGATTTC	AACTCGGGTG	CGGCCGAGAT	GACCGGCTTC	CTCACCGGTC	TCAGCGGCCG
CAGCTAAAGT	TTGAGCCAC	GGGGGCTCTA	CTGGCCGAAG	GAGTGGCCAG	AGTCGCCGGC
7630	7640	7650	7660	7670	7680
CTTCGTGCGC	CCCGGTCCAT	CGGGCGCGCC	GACCGCGCGC	CGGGCCGGATG	TGTTGCGAC
GAAGCAGCGG	GGGCCAGGT	GGCGCGCGG	CTGCGCGCCG	GCCGGCCTAC	ACAAACGGCTG
7690	7700	7710	7720	7730	7740
GGGGCGCAAT	TTCTACTCGG	TCGACAGCGC	CGCCGTCGCG	ACGCCGGCGG	CITACGAGCT
CCCCCGCTTA	AAGATGAGCC	AGCTGTCGCG	GGGGCACCGC	TGCGGCCGCC	GAATGCTCGA

FIGURE 43.7

PL. 119 | 15 |

7750	7760	7770	7780	7790	7800
TGGCAAGAAA	TCGGCCGAGC	TTCTGATCCG	CCGCTACCTG	CAGGACCATG	GCGAATGGCC
ACCGTTCTT	AGCCGGCTCG	AAGACTAGGC	GGCGATGGAC	GTCTGGTAC	CGCTTACCGG
7810	7820	7830	7840	7850	7860
GTCCCTCCCTT	GGCCTGACCG	CCTGGGGCAC	GGCGAACATG	CGCACCGGGCG	GCGACGACAT
CAGGAGGAAA	CCGGACTGGC	GGACCCCGTG	CCGCTTGAC	CGTGGCCGC	CGCTGCTGTA
7870	7880	7890	7900	7910	7920
CGCCCCAGGCC	CTGGCGCTGA	TCGGCGCCAA	GCCCACCTGG	GACATGGTCT	CTCGCCGGGT
CGGGGGTCCGG	GACCCGCACT	AGCCGCGGTT	CGGGTGGACC	CTGTACCAGA	GAGCGGCCCA
7930	7940	7950	7960	7970	7980
GATGGGCTAC	GAGATCGTGC	CGCTCGCAGT	CCTCGGGCCG	CCACCGTCTG	ACGTGACCTT
CTACCCGATG	CTCTAGCAGC	GGAGCGTCA	GGAGCCGGCG	GGTGCAGCAGC	TGCACTGGAA
7990	8000	8010	8020	8030	8040
GCGCATTTCC	GGCTTCITCC	GGATGCCTT	CCCGGACCG	ATCGCGCTCT	TCGACAAGGC
CGCGTAAAGG	CCGAAGAAGG	CGCTACGGAA	GGGCCCTGGTC	TAGCGCGAGA	AGCTGTTCCG
8050	8060	8070	8080	8090	8100
GATCCCGCGCC	GTCCGCGCTGG	AGGAAGACGA	TGCGGACAC	ATGATCGCCG	CACCGATGCG
CTAGGGCGGG	CAGCGCGACC	TCCTTCTGCT	ACGGCTGTG	TACTAGCGGC	GTGCGTACGC
8110	8120	8130	8140	8150	8160
GGCGGAAAGC	CGGGGGCTGG	AGGCCGAAGG	CGTGGAAGCC	GCCGAGGCCG	CGCGTCGCGC
CGCGCTTTCG	GGCGCCGACC	TCCGGCTTCC	GCACCTTCGG	CGGCTCCGGC	GCGCAGCGCG
8170	8180	8190	8200	8210	8220
CTCCCTACCGC	GTCTTTGGCG	CAAAGCCCGG	TGCGCTATGGC	GCCGCCCTGC	AGGCCTGTAT
GAGGATGGCG	CAGAAACCGC	GTTCGGGGCC	ACGGATAACCG	CGGGGGACG	TCCGGACTA
8230	8240	8250	8260	8270	8280
CGACGAGAAG	GGCTGGGAAA	CAAAGCAGA	TCTCGCCAG	GCCTATCTTA	CCTGGGGCGC
GCTGCTCTTC	CCGACCCCTT	GGTTTCGTCT	AGAGCGCTC	CGGATAGAAT	GGACCCCGCG
8290	8300	8310	8320	8330	8340
CTATGCCTAT	GGCGCCGGCG	AGGAGGGCAA	GGCGGAGCGC	GATCTTTTCG	AGGAGCGCCT
GATACGGATA	CCGCGGCCGC	TCCTCCCGTT	CCGGCTCCGG	CTAGAAAAGC	TCCCTCGCGGA
8350	8360	8370	8380	8390	8400
GCGCACGATA	GAGGCGGTGG	TGCAGAACCA	GGACAACCGC	GAGCACGATC	TGCTCGACAG
CGCGTGTAT	CTCCGCCACC	ACGTCTTGGT	CCTGTTGGCG	CTCGTGCTAG	ACGAGCTGTC
8410	8420	8430	8440	8450	8460
CGACGACTAC	TACCAAGTTCG	AAGGCGGCAT	GAGCGCTGCC	GCCGAACAGC	TCGGCGGTCA
GCTGCTGTG	ATGGTCAAGC	TCACGCCGTA	CTCGCGACGG	CGGCTTGTG	AGCCGCCAGT
8470	8480	8490	8500	8510	8520
CCGTCCGGCG	ATCTACCAACA	ACGACCAATTG	CCGTCCGGAA	AAGCTGTGA	TCCGGTCGCT
GGCAGGGCCGC	TAGATGGTGT	TGCTGGTAAG	GGCAGGCCCT	TTCGGACACT	AGGCCAGCGA
8530	8540	8550	8560	8570	8580
CGAAGAAAGAG	ATCGGCCGCG	TGGTCCGGGC	CCGCGTGTG	AATCCAAGT	GGATCGATGG
GCTTCTTCTC	TAGCCGGCGC	ACCAGGCCCG	GGCGCAGCAG	TTAGGGTTCA	CCTAGCTACC
8590	8600	8610	8620	8630	8640
CGTCATGCGC	CACGGATAACA	ACGGCGCCCT	CGAGATCGCT	GCCACGGTGC	ACTACATGTT
GCAGTACGCG	GTGCCTATGT	CCCCGGCGAA	GCTCTAGCGA	CGGTGCCAGC	TGATGTACAA
8650	8660	8670	8680	8690	8700
CGCCTTGCC	GCGACCAACGG	GTGCGGTGCG	CGACCATCAT	TTCGAGGCCG	CTTATCAGGC
GCGGAAACGG	CGCTGGTGCC	CAACGCCACGC	GCTGGTAGTA	AAGCTCCGGC	GAATAGTCCG

FIGURE 43.8

PL. 120 (15)

8710	8720	8730	8740	8750	8760
GTTCAATTGTC	GACGAGCGCG	TGGCTGACTT	CATGCCGAC	AAGAACCCGG	CCGCCTTTCG
CAAGTAACAG	CTGCTCGCG	ACCGACTGAA	GTACCGCCTG	TTCTGGGCC	GGCGGAAACG
8770	8780	8790	8800	8810	8820
CGAGCTTGC	GAACGCCCTGC	TTGAAGCAAT	CGACCCAAT	CTCTGGACGC	CGCGCTCGAA
GCTCGAACGG	CTTGCAGACG	AACTTCGTTA	GCTGGCGTTA	GAGACCTGCG	GCGCGAGCTT
8830	8840	8850	8860	8870	8880
TTCGGGCGCG	TTTGAACTTG	CCGGCATCGG	CACCGCAGCA	ACCCGGCTTC	GTGCCGGCAA
AAGCCCGGCC	AAACTTGAAC	GGCGTAGGCC	GTGCCGTGCG	TGGGCCGAAG	CACGGCCGTT
8890	8900	8910	8920	8930	8940
TGAATAGAGC	GGTTCCGGGC	TGGCGGTAT	CCGTCCGGAA	TTGCTTGGAA	ACAAAAGACCT
ACTTATCTCG	CCAAGGCCCC	ACCGCCAATA	GGCAGGCCCT	AACGAACCTT	TGTTTCTGGA
8950	8960	8970	8980	8990	9000
GGTTCCGTT	CGCTGCTCAG	TGAAGTGCAG	AAAGGAACCG	AAGCGGGACG	AGGGCGCTCG
CCAAGGCAAA	GGGACCGAGTC	ACITCACCGT	TTTCCCTGGC	TTCGCCCTG	TCCCAGACAC
9010	9020	9030	9040	9050	9060
CCCATCCCGA	ACTTGAGAAC	TGAGGGAGTG	ATCATGAGCG	ACGAGACGAC	AGTAGGCGGC
GGGTAGGGCT	TGAACCTTGT	ACTCCCTCAC	TAGTACTCGC	TGCTCTGCTG	TCATCCGCCG
9070	9080	9090	9100	9110	9120
GAAGCCCCGG	CCGAGAAGGA	CGATGCCCGC	CACGCCATGA	AGATGGCGAA	GAAGAAGGCA
CTTCGGGGCC	GGCTCTTCCT	GCTACGGGCG	GTGCGGACT	TCTACCGCTT	CTTCTTCCGT
9130	9140	9150	9160	9170	9180
GCACCGCGAA	AGATCATGGC	GACGAAGACC	GACGAGAAGG	GTCTGATCAT	CGTCAACACCC
CGTGCCTTT	TCTAGTACCG	CTGCTTCTGG	CTGCTCTTCC	CAGACTAGTA	CGAGTTGTGG
9190	9200	9210	9220	9230	9240
GGCAAAGGCA	AGGGCAAGTC	GACCGCCGGC	TTCCGGCATGA	TCTTCCGCCA	TATCGCCCAC
CCGTTTCCGT	TCCCAGTCAG	CTGGCGGCCG	AAGCCGTACT	AGAAGGCGGT	ATAGCGGGTG
9250	9260	9270	9280	9290	9300
GGCATGCCCT	GGCCCGTCGT	GCAGTTCATC	AAGGGTGCAG	TGGCAACCGG	CGAGCGCGAG
CCGTACGGGA	CGCGGCGAGCA	CGTCAAGTAG	TTCCCACGCT	ACCGTTGGCC	GCTCGCGCTC
9310	9320	9330	9340	9350	9360
TTGATCGAGA	AGCAATTGCG	CGATGCTCGC	CAGTTCTACA	CGCTCGCGA	GGGCTTCACC
AACTAGCTCT	TCGTAAGGCC	GCTACAGACG	GTCAAGATGT	GCGAGCCGCT	CCCCAAGTGG
9370	9380	9390	9400	9410	9420
TGGGAAACCGC	AGGATCGCGC	CCGCGATGTT	GCGATGGCTG	AAAAGGCCTG	GGAGAAGGCG
ACCCCTTGC	TECTAGCGCG	GGCGCTACAA	CGCTACCGAC	TTTCCGGAC	CCCTCTTCCGC
9430	9440	9450	9460	9470	9480
AAGGAACGTG	TCCGTGACGA	GCGCAACTCG	ATGGTGTGCTG	TCGACGAGAT	CAACATTGCT
TTCCTTGACT	AGGCACTGCT	CGCGTTGAGC	TACCAACGACG	AGCTGCTCTA	GTGTAACGA
9490	9500	9510	9520	9530	9540
CTGCGCTACG	ACTACATCGA	CGTCGCCGAA	GTGGTGCCTG	TCCTGAAGGA	AGAAAAGCCG
GACGCGATGC	TGATGTAGCT	GCAGCGGCTT	CACCAACGCGA	AGGACTTCCT	TCTTTTCCGGC
9550	9560	9570	9580	9590	9600
CACATGACGC	ATGTGGTGCT	CACCGGCCGC	AACCGCGAAAG	AAGACCTGAT	CGAAGTCGCC
GTGTACTGCG	TACACCACGA	GTGGCCGGCG	TTGCGCTTTC	TTCTGGACTA	GCITCAGCGG
9610	9620	9630	9640	9650	9660
GATCTCGTC	CTGAGATGGA	GTGTGATCAAG	CATCCGTTCC	GTTCGGGCAT	CAAGGCGCAG
CTAGAGCAGT	GACTCTACCT	CGACTAGTC	GTAGGCAAGG	CAAGGCCGTA	GTTCGGCGTC
9670	9680	9690	9700	9710	9720

FIGURE 43.9

PL. 121 / 15)

CAGGGCGTGG AGTTCTGATG AGCCAGAGCT GGCAGTCTG GGCGCTGCTT TCGGCCGCC  
 GTCCCCCACC TCAAGACTAC TCGGTCTCGA CCGTCAGAC CGCGACGAA AGCCGGCGGA

9730	9740	9750	9760	9770	9780
TCGCTGCGCT	CACCGCGGTG	TFTGCCAAGG	TCGGGGTTGC	GCAGATCAAC	TCCGACTTCG
AGCGACGCGA	GTGCGGCCAC	AAACGGTTCC	AGCCCCAACG	CGTCTAGTTG	AGGCTGAAGC

9790	9800	9810	9820	9830	9840
CAACGCTGAT	CCGCACCGTC	GTCATCCTCT	GGGTGATCGC	CGCCCATCGT	GGGGCGACAG
GTGCGACTA	GGCGTGGCAG	CAGTAGGAGA	CGCACTAGCG	GGGGTAGCAC	CGCCGCTGTC

9850	9860	9870	9880	9890	9900
GGCAGTGGCA	GAAGCCATCG	GAAATCCCGG	GGCGCACCTG	GCTGTTCTG	GCGCTGTCAG
CCGTCAACCGT	CTTCGGTAGC	CTTTAGGGC	GGCGTGGAC	CGACAAGGAC	CGCGACAGTC

9910	9920	9930	9940	9950	9960
GGCTTGCAC	TGGCGCTTCC	TGGCTTGCCT	ATTCGGCGC	GCTGAAGCTC	GGCGACGCCG
CCGAACGCTG	ACCGCGAAGG	ACCGAACGGA	TAAAGCCCGG	CGACTTCGAG	CCGCTGCGGC

9970	9980	9990	10000	10010	10020
CCCGCGTGGC	GCCGCTCGAC	AAGCTCTCGA	TCGTCATGGT	CCCGATCTTC	GGCGTGCTCT
GGGCGCACCG	GGCGGAGCTG	TTCGAGAGCT	AGCAGTACCA	GGCGTAGAAG	CCGCACGAGA

10030	10040	10050	10060	10070	10080
TCCTCGGTGA	AAAGCTCAAC	CTGATGAAC	GGCTGGCGT	CCCGTTCAATT	GGCGCCGGGG
AGGAGCCACT	TTTCGAGTTG	GACTACTTGA	CCGAGCCGCA	GGGAAAGTAA	GGGCGGCCCC

10090	10100	10110	10120	10130	10140
CGCTGTGCT	GGCGGTGTT	TGAGCGCGCC	TGCTCTGGTG	CCTGTTCACT	GAATGCTCGC
GCGACAACGA	CCGCCACAAA	ACTCGCGGG	ACGAGACCAC	GGACAAGTGA	CTTACGAGCG

10150	10160	10170	10180	10190	10200
CTCAATCAAT	CCGTAATCCC	GACACATGCA	GTGGTTGTGA	CGAGCGGGAG	GACGGCATGC
GAGTTAGTA	GGCATTTAGGG	CTGTGTACGT	CACCAACACT	GCTCGCCCTC	CTGCGGTACCG

10210	10220	10230	10240	10250	10260
AGATTGAAGG	CAATTGGAGC	GAGCGCTTC	CTGATCCGTC	GGGCCACGTC	GGCGAGTTCG
TCTAACCTCC	GTAAACCTCG	CTCGCGGAAG	GACTAGGCAG	CCGGGTGCAG	CGCGTCAAGC

10270	10280	10290	10300	10310	10320
GCAGACCGCTG	GAAGCGTCGC	AGCCTGAGGG	TGAGCCCTGC	TTCAGACCCA	CCGGCGGGACA
CGTCTGCGAC	CTTCGCAAGCG	TEGGACTCCC	ACTCGGACG	AAAGTCTGGGT	GGCCGCGCTGT

10330	10340	10350	10360	10370	10380
CGCCTGCAAT	AGGCACCGTA	GGCGTCGCGG	AAGACCTTGG	CGAGGGGGT	TTCCCTCCATG
GCGGACGTTA	TCCGTGGCAT	CCGCAGCGGC	TTCTGGAACC	GCTCCACCCA	AAGGAGGTAC

10390	10400	10410	10420	10430	10440
CGGATCTGGT	AGGAAATCGA	GATCCAGGGC	GAGAGCCCA	GCGCCACCGA	GATGACGTTG
GCCTAGACCA	TCCCTTAGCT	CTAGTCCGC	CTCTCGCGT	CGCGGTGGCT	CTACTGCAAC

10450	10460	10470	10480	10490	10500
GGCACCGCCA	TCACCGTGC	GATCAGCGCG	GTCACCATGC	CGACATAGAT	GGGGTTGCC
CCGTGGCGGT	AGTGGCACGG	CTAGTCCGC	CAGTGGTAGC	GCTGTATCTA	GCCCAACCGCG

10510	10520	10530	10540	10550	10560
GAGAAGGCAT	AGAGGCCTGA	GGTCACAAGC	GGCGCGTCCT	GCTTTTCAGG	GATGCCGATC
CTCTTCCGTA	TCTCCGGACT	CCAGTGTTCG	CCGCGCAGGA	CGAAAAGTCC	CTACGGCTAG

10570	10580	10590	10600	10610	10620
TTCCAGGAAT	GACGCATCGC	CCATTGCGAC	AGCATCGTC	GCCCGCCGCC	GAGCGTCATC
AAGGTCCCTA	CTGCGTAGCG	GGTAACGCTG	TGTTAGCACT	CGGGCGGGCG	CTCGCAGTAG

10630	10640	10650	10660	10670	10680
AGCGCCAGGC	CGACGGCGTG	AAGGATGGGC	GTGTCAGCG	CCGGGATCCG	GCCGAGGGCA

FIGURE 43.10

PL. 122 / 15)

TCGCGGGTCCG GCCTGCCGAC TTCTTACCCG CACAGCTCGC GGCCCTAGGC CGGCTCCCGT  
 10690 10700 10710 10720 10730 10740  
 GCATCGACGG AGGCCGGGAG CATGGCGACC GCCAGCAGGT GGATCACCAAG CGCTCGACG  
 CGTAGCTGCC TCCGGCCCTC GTACCGCTGG CGGTGGTCCA CCTAGTGGTC GCGACGCTGC  
 10750 10760 10770 10780 10790 10800  
 ATCAGGCGGA AAAGCCTGCC CGCAAACCT CGCCATCGT CGCCATAGGT TAGCACGACC  
 TAGTCCGCCT TTTCGGACGG GCGTTGGGA AGGCGTAGCA GCGGTATCCA ATCGTGTGG  
 10810 10820 10830 10840 10850 10860  
 GGCAGCGGC CGGAATTGAC GCGGGCGAGG ATCGCCAGCG CGAGCGTGGGA CAATCCCACG  
 CCGCTCGCC CGCTAACGTG CGCCGCCTCC TAGCGTCGC GCTCGCACCT GTTAGGGTGC  
 10870 10880 10890 10900 10910 10920  
 ACGAGCATCA GGATGGTGGG AAGGGTGGTG GACATGGAAA CCTCTGGAGC GAGCTGACAA  
 TGCTCGTAGT CCTACCAACCC TTCCCCACAC CTGTACCTT GGAGACCTCG CTCGACTGTT  
 10930 10940 10950 10960 10970 10980  
 GACAGGAGCG CACGACGGGT AGGCGGCCA TATGAGCGTC TACCCGGCGA AGCATTCTGA  
 CTGTCCCTCGC GTGCTGCCA TCCGCCGGT ATACTCGCAG ATGGGCCGCT TCGTAAGACT  
 10990 11000 11010 11020 11030 11040  
 TCACCTTGCA ATCTCTAGTA ACTAGAGGTT CAAGCGTCGG ACCTGTCCGA CTTTCGTCGT  
 AGTGGAACGT TAGAGATCAT TGATCTCAA GTTCCGAGCC TGGACAGGCT GAAAGCAGCA  
 11050 11060 11070 11080 11090 11100  
 GGTTACCGGA TCTTATTGCC AAGCCTTGGGA GGCTGTCACT GTCGCCCCCG CCCTGTGGGA  
 CCAATGGCCT AGAATAACGG TTGCAACCT CCGACAGTAG CAGCGGGGGC GGCACAGCCT  
 11110 11120 11130 11140 11150 11160  
 AGGTGGCAA AATTGCTCTC TTGACGGCTG CTCCCTCCGT CGAGCGATTG CATAGGCAGG  
 TCCAGCCGT TTAAGCAGAG AACTGCCGAC GAGGAAGGCA GCTCGCTAAC GTATCCGTCC  
 11170 11180 11190 11200 11210 11220  
 AGGCCGCACC CATGTTAGAC CGTCGACAGG CTAAATACGG GTGAACCTTG AAGAATACTC  
 TCCGGCGTGG GTACAATCTG GCAGCTGTCC GATTATGCC CACTTGGAAC TTCTTATGAG  
 11230 11240 11250 11260 11270 11280  
 TCAGAGCTGC GTTGGTGTGC GCATCGGTCT TGCTGTCTT GTCATCAGGT GTGGCGGGGC  
 AGTCTCGACG CCAACACAG CGTAGCCAGA ACGACAAGAA CAGTAGTCCA CACCGCCCCG  
 11290 11300 11310 11320 11330 11340  
 AGGCGCAAAC CGTGAAGAGC GGGGCGTCAC GAGCTCAAGA AACGACGACC ACCCAGAAGG  
 TCCCGTTTG GCACCTCTCG CCCCCGAGTG CTCGAGTTCT TTGCTGCTGG TGGGTCTTCC  
 11350 11360 11370 11380 11390 11400  
 CGAAACCGAA AACTAAAACG ACAGCGCAAGC AAAGGGCTGC GGATGAAGCC AAGGCCAAGG  
 GCTTGGCTT TTGATTTGC TGCGCGTTGG TTTCCCAGC CCTACTTCGG TTCCGGTTCC  
 11410 11420 11430 11440 11450 11460  
 CGCTCGCCGA AGCGCGCCGT CCACGGATTG GCAAGACGCG GGAGAGCGAA TGCAGCTATG  
 GCGAGCGGCT TCGCGCGCA GGTGCTAAA CGTTCTGCGC CCTCTCGCTT ACGTCGATAC  
 11470 11480 11490 11500 11510 11520  
 GCGCAGGTCC GGTGGAGAG CAGTGCTCGT GCTGGTCGAA ATCCGGTGC CGTGCATCTG  
 CGCGTCCAGG CCAGCCTCTC GTACGAGCA CGACCAAGCTT TAGGCCACGC GGACTAGAAC  
 11530 11540 11550 11560 11570 11580  
 GCATAACTGT CAGGGCTTGA CGGCCCGCA CCTTCGGCGG GGCAGGCAAG CGTGCCTCGC  
 CGTATTGACA GTCCGCAACT GGCAGGGCGT GGAAGCGCGC CGTCCGTTTC GCACGCAGCG  
 11590 11600 11610 11620 11630 11640  
 TCGAAGCGAC GCCTGACGCG ATAGAAATCA CGGGTCCGCCT GGTTGTTCT GAAAGCTTGG  
 AGCTTCGCTG CGGACTGCGC TATCTTGTG GCCCAGCGGA CCAAGCAAGA CTTTCGAACC

FIGURE 43.11

PL. 123 / 15 /

11650	11660	11670	11680	11690	11700
GATGGGTTT	AGGTGATGGA	AGCCGGCGTT	GAACGCAGAA	TAATGATCGA	TCTCGAGAAC
CTAACCCAAA	TCCACTACCT	TCGGCCGCAA	CTTGCCTTTT	ATTACTAGCT	AGAGCTCTTG
11710	11720	11730	11740	11750	11760
AGCCGCGCTCC	AGTTTGCACAC	CCGAGCACAC	GGCGAACAGA	AGCGTAAGTA	TGACGGTCGG
TCGCCGAGG	TCAAAACGTTG	GGCTCGTGTG	CCGCTTGTCT	TCGCATTCAT	ACTGCCAGCC
11770	11780	11790	11800	11810	11820
CCCTATATCG	TTCATCCGAT	TGCGGTGGCG	GAGATTGTTG	GAAGCGTGCC	CCATACGCC
GGGATATAGC	AAGTAGGCTA	ACGCCACCAG	CTCTAACAAAG	CTTCGACCGG	GGTATGCGGG
11830	11840	11850	11860	11870	11880
GAAATGATCG	CCGCAGCGCT	GCTTCACGAT	ACGGTCGAAG	ATACCGACGC	GACGCTGCTG
CTTTACTAGC	GGCGTCGCGA	CGAAGTGCCTA	TGCCAGCTTC	TATGGCTGCG	CTGCGACGAC
11890	11900	11910	11920	11930	11940
GAGATCAAGG	AAGCGTTCCG	CCCCAAGGTC	GCAACACTGG	TTGCGTGGCT	CACCGACATA
CTCTAGTTCC	TTCGCAAGCC	GGGGTTCCAG	CGTTGTGACC	AACGCACCGA	GTGGCTGTAT
11950	11960	11970	11980	11990	12000
TCCACTCCGT	TCCACGGCAA	CCGACAGGTG	CGCAAGGAAC	TGGATCGCCA	GCACCTCGCA
AGGTGAGGCA	AGGTGCCGTT	GGCTGTCCAC	GCCTTCCTTG	ACCTAGCGGT	CGTGGAGCGT
12010	12020	12030	12040	12050	12060
TCGGCGCCCG	CCGCGGGCGAA	AACCGTCAAG	CTCGCCGACC	TGATCGACAA	TGCGATAGCG
AGCCCGGGGC	GGCGCCGCTT	TTGGCAGTTC	GAAGGSCCTGG	ACTAGCTGTT	ACGCTATCGC
12070	12080	12090	12100	12110	12120
ATCAAAGCCG	GCGATCCGAA	TTTCTGGAAA	GTGTTGGCG	CCGAGATGAA	ACGCTTGTG
TAGTTCGGC	CGCTAGGCTT	AAAGACCTTT	CACAAGCCGC	GGCTCTACTT	TGCGAACGAC
12130	12140	12150	12160	12170	12180
GAGGCTTCTGG	GCGACGGCGA	CGAGACCCCTT	CTCGCAAAGG	CCCCTGCATT	AGGCCGGAA
CTCCAGAAC	CGCTGCCGCT	GCTCTGGGAA	GAGGTTTCC	GGGCACGTAA	TCGGGCCCTT
12190	12200	12210	12220	12230	12240
TGAGAGTGCC	GCCGTTTATC	GGCAAGCATG	TCTGTGCCAT	GTGACCCGG	TCAACCGTC
ACTCTCACGG	CGGCAAATAG	CCGTTCGTAC	AGACACGGTA	CAGCTGGGCC	AGTTGGCCAG
12250	12260	12270	12280	12290	12300
ATCCAAGATC	GCAGAACCGA	CATGCATTG	CGGTTTGCC	CGCCGGTGTG	GCCCAGCCAC
TAGGTCTAG	CGTCTTGCTT	GTACGTAAC	GCCAAAACGG	GCGGCCACAC	CGGGTCGGTG
12310	12320	12330	12340	12350	12360
GCCTCACAGG	CTGCGCGGTT	GCGGCCGTTA	GGACAGCGA	GAATTGCGC	ACCGCGCCGC
CGGAGGTGTC	GACCGCGCAA	CGCCGGCAAT	CCTGTGCGT	CTTAAACGGC	TGGCGCGCG
12370	12380	12390	12400	12410	12420
GCCTCAATGC	CCCAGCCAGA	TCCGCAAGGG	ATGCGTCGGA	TCTGCGAGCA	GGCGGATCGC
CGGAGTTACG	GGGTCGGTCT	AGGGCTTCCC	TACGCAACCT	AGACGCTCGT	CGGCCTAGCG
12430	12440	12450	12460	12470	12480
GAGCGCGATC	GAGACGATGA	CGAGCAGCGG	CTTGATGATC	TTGGCGCCCT	TGGCGATGCC
CTCGCGCTAG	CTCTGCTACT	GCTCGTCGCC	GAACACTAG	AACCGCGGGA	ACCGGTACCG
12490	12500	12510	12520	12530	12540
ATAGCGCGAG	CCGACCTGGG	GGCGGAGGAA	CTGGCCGAGG	CCCACATCAACA	GGCCGACCTT
TATCGCGCTC	GGCTGGACCC	GGGGCTCCTT	GAACGGCTCC	GGGTAGTTGT	CGGGCTGGAA
12550	12560	12570	12580	12590	12600
CCAGAGAACG	GCGCCGAAGA	AGAGGAAGAC	GCCGAAGGCG	CCGACGTTGG	AGCCAAAGTT
GGTCTCTTGC	CGCGGCTTCT	TCTCCCTCTG	GGGCTTCCGC	GGCTGCAACC	TGGTTTCAA

FIGURE 43.12

FEUILLE DE REMplacement

PL. 124 / 15 /

12610	12620	12630	12640	12650	12660
GAGGAACCTTC	GTGTGCGCCG	TGCCCCCAA	CACGCCGAAG	CCGGCGAGGG	TAACGAAGCC
CTCCTTGAAG	CACACGECCC	AGCGGAAGTT	GTGCGGCTTC	GGCCGCTCCC	ATTGCTTCGG
12670	12680	12690	12700	12710	12720
GAGCATGAAAG	AACCGAGCCGG	TGCCCCGGGC	GAAGACGCCG	TCATAAAAAGC	CGATTTAGCGG
CTCGTACTTC	TTGCTCGGCC	ACGGGCCCCGG	CTTCTGCGGC	AGTATTTTCG	GCTAATCGCC
12730	12740	12750	12760	12770	12780
CACCAGTGTC	AGCGTGAAAGA	CGAAGGGGGT	GACGCCGCTG	TGCTGGTCGA	CGTCGCCCAT
GTGGTCACAG	TCGGCACTTCT	GCTTCCCCCA	CTGCGCCGAC	ACGACCAGCT	GCAGCGGGTA
12790	12800	12810	12820	12830	12840
GTTCGGCTTC	AGGGCGAAAT	AAAGCGCAAT	GGCGATCAGC	AGAAAGGGCA	GGATCGCCCTT
CAAGCGGAAG	TCCGGCTTTA	TTTCGCGITA	CCGCTAGTCG	TCTTCCCGT	CCTAGCGGAA
12850	12860	12870	12880	12890	12900
CAGCACGTG	CCGGGAACGA	TGGTTGCGAG	CAGGGCCCG	AGCACGGCGC	CGGCGGCCGA
GTCGTGCAGC	GGCCCTTGCT	ACCAACGCTC	GTCCCCGCGC	TGTTGCCGCG	GCCGCCGGCT
12910	12920	12930	12940	12950	12960
CATCAGCGCC	ATCGGCACT	GCTCTTTCAG	GTTCACGTGG	CCGCGCCGGG	CATAGGACAG
GTAGTCCGG	TAGCCGTCGA	CGAGAAAGTC	CAAGTGCACC	GGCGCGGCC	GTATCCTGTC
12970	12980	12990	13000	13010	13020
CGTGGCCGAG	CCGGAGCCGA	ACAATCCCTG	CAGCTTGTG	GTGCCGAGCG	TCTGCAAGGG
GCACC GGCTC	GGCCTCGGCT	TGTTAGGGAC	GTCGAACAC	CACGGCTCGC	AGACGTTCCC
13030	13040	13050	13060	13070	13080
CGGGATGCC	GCAATGAGCA	TGGCCGGAAT	GGTGATCATG	CCACCGCCGC	CGGCATCGA
GCCCTACGGG	CGTTACTCGT	ACCGGCCCTTA	CCACTAGTAC	GGTGGCGGCG	GCCGCTAGCT
13090	13100	13110	13120	13130	13140
ATCGATGAAG	CCTGCGATGA	AGGCAGCGAC	GAACAGGAAG	GGGAGCAGGT	GGAAAGGCGAG
TAGCTACTTC	GGACGCTACT	TGCGCCGCTG	CTTGTCTTC	CGCTCGTCCA	CCTTCCGCTC

ATCT  
TAGA

FIGURE 43.13

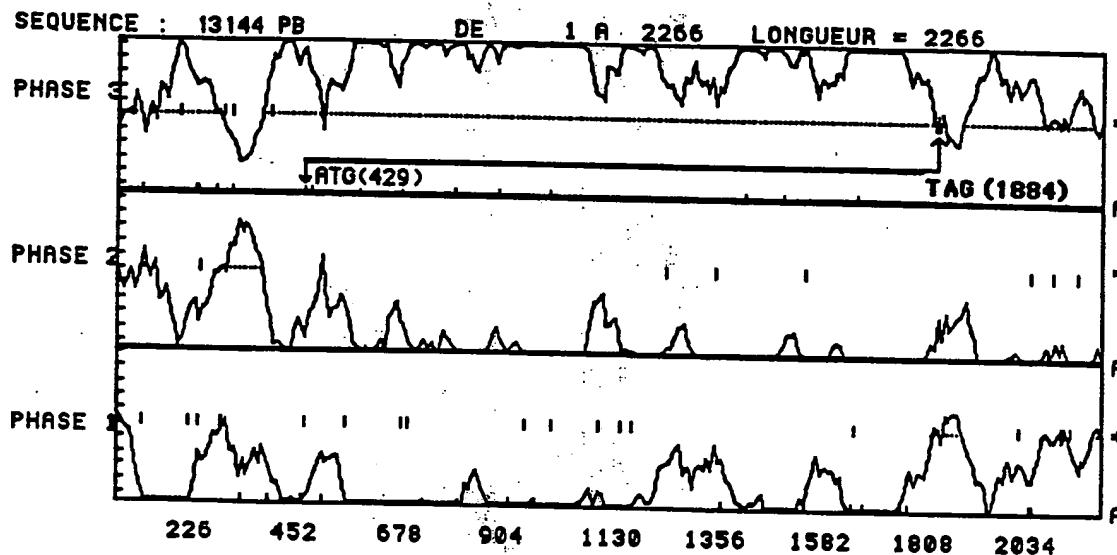
125 / 151

## Carte de restriction de la sequence de 13144 pb

ApaI	642,
EcoRI	8818,
HindIII	11633,
MluI	7963,
NdeI	10950,
PvuII	12918,
SfiI	3133,
SphI	99,
BglII	8248, 13139,
KpnI	2315, 6300,
NotI	5526, 7615,
SmaI	1322, 9868,
SspI	4843, 6968,
XbaI	9313, 12091,
AatII	1033, 9503, 12773,
AflIII	550, 7963, 8634,
BalI	2107, 6236, 12473,
BamHI	2266, 5416, 10664,
BspMII	5002, 8494, 8914,
EcoRV	4263, 4605, 5137,
NcoI	6318, 7786, 12474,
NsiI	3467, 5064, 12266,
PflMI	7870, 10718, 11065,
XbaI	1512, 4171, 11692,
Apal	1928, 3138, 3386, 8551,
AsuII	784, 5670, 8418, 11799,
FspI	784, 5670, 8418, 11799,
MaeI	1883, 2647, 10995, 11002,
NruI	1827, 3794, 10002, 12419,
SauI	852, 7001, 10284, 10517,
BstEII	995, 3642, 8456, 10470, 11041,
Eco47III	6954, 7209, 8434, 10731, 11837,
SacI	5, 4109, 4694, 5169, 11315,
StuI	204, 4081, 8261, 9406, 10515,
BstXI	761, 2982, 3612, 6031, 6232, 9102,
SacII	932, 1025, 2096, 3537, 5184, 12014,
SphI	966, 2740, 5360, 8098, 9246, 10199,
BclI	2992, 4016, 9029, 9164, 9623, 10978, 13053,
RsaI	101, 1201, 1918, 2313, 4881, 6298, 6856,
Tth111I	1821, 2424, 4351, 7361, 7904, 12227, 12697,
PstI	613, 3989, 5832, 5952, 7260, 7782, 8211, 12992,
ClaI	1351, 3596, 4469, 4724, 5748, 6618, 8574, 11687,
	13082,
FspI	1363, 1551, 1653, 5219, 7841, 7982, 8342, 9760,
	11971,
HinfI	1137, 2564, 2592, 3025, 5667, 5927, 6467, 6923,
	13079,
StyI	2488, 3396, 5116, 6105, 6318, 7786, 9745, 10355,
	11389, 11395, 11903, 12468, 12474,
DdeI	852, 1875, 3373, 3586, 6311, 7001, 7010, 7610,
	8956, 9020, 9611, 10284, 10517, 11220,
Nsp7524I	554, 966, 2394, 2740, 5360, 7840, 8098, 8638,
	9246, 9553, 10168, 10199, 12210, 12266,
PvuI	26, 1853, 2453, 4403, 4703, 4728, 5091, 5112,
	5178, 6717, 7269, 9991, 12429, 13077,
AvaI	975, 1320, 1503, 1512, 3131, 3231, 3709, 3766,
	4171, 4212, 7224, 7573, 9866, 11692, 11720,
BanII	5, 496, 1723, 1928, 2254, 3138, 3386, 4109,
	4694, 5169, 6207, 6282, 8551, 10296, 11315,
SalI	83, 1296, 2418, 4045, 4303, 5258, 6959, 7700,
	7967, 8627, 8708, 9198, 11182, 12221, 12766,
XbaII	2266, 3920, 5416, 5688, 6943, 7020, 7140, 8248,
	10382, 10400, 10664, 11048, 12378, 12398, 13139,

126 / 151

## PHASE OUVERTE 22



## PHASE OUVERTE 23

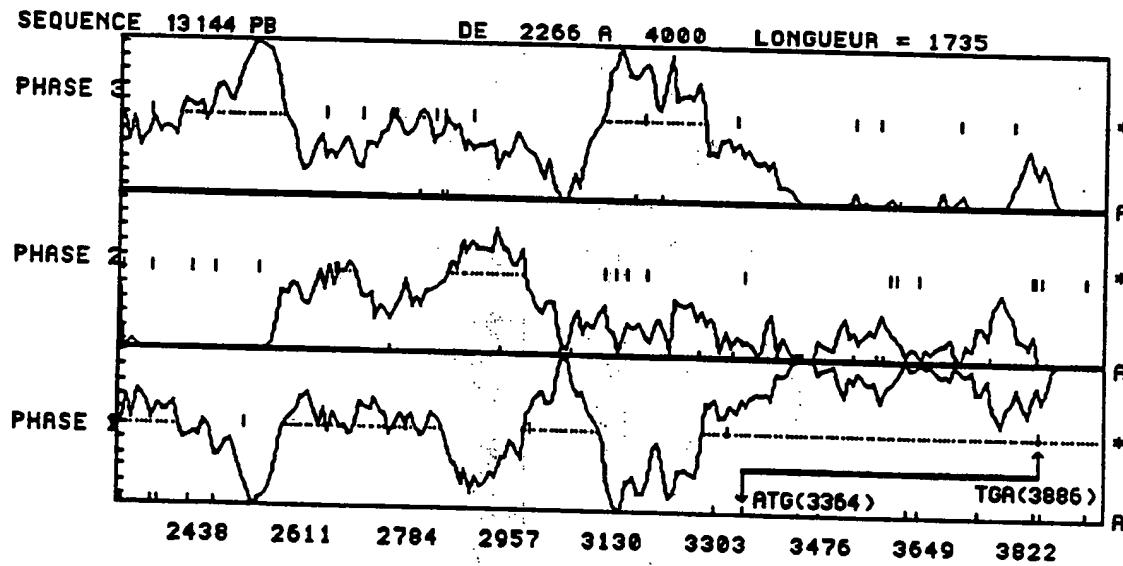


FIGURE 45

127/451

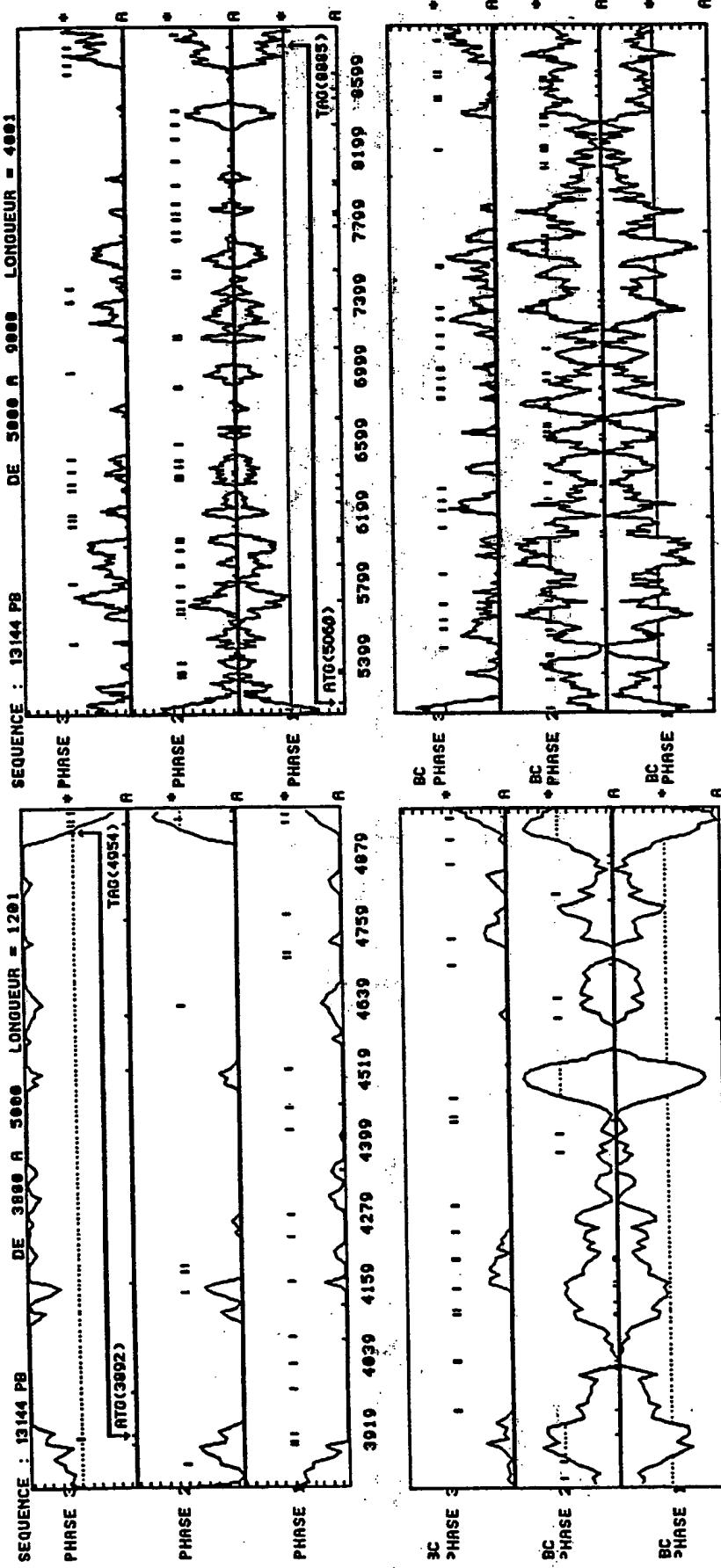
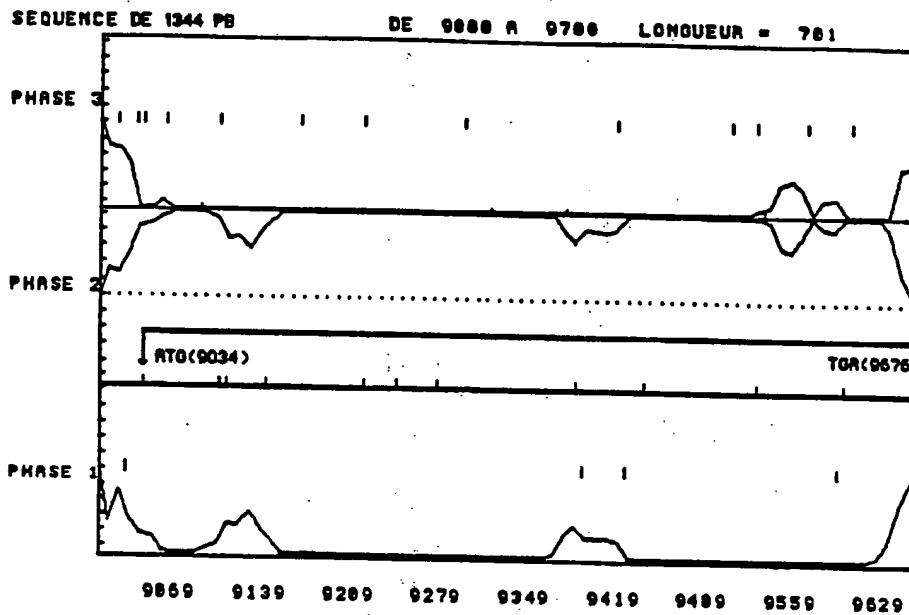


Fig.45.1

PHASE OUVERTE 25

PHASE OUVERTE 24

128 / 151



SEQUENCE DE 13144 PB DE 9888 A 13144 LONGUEUR = 3545

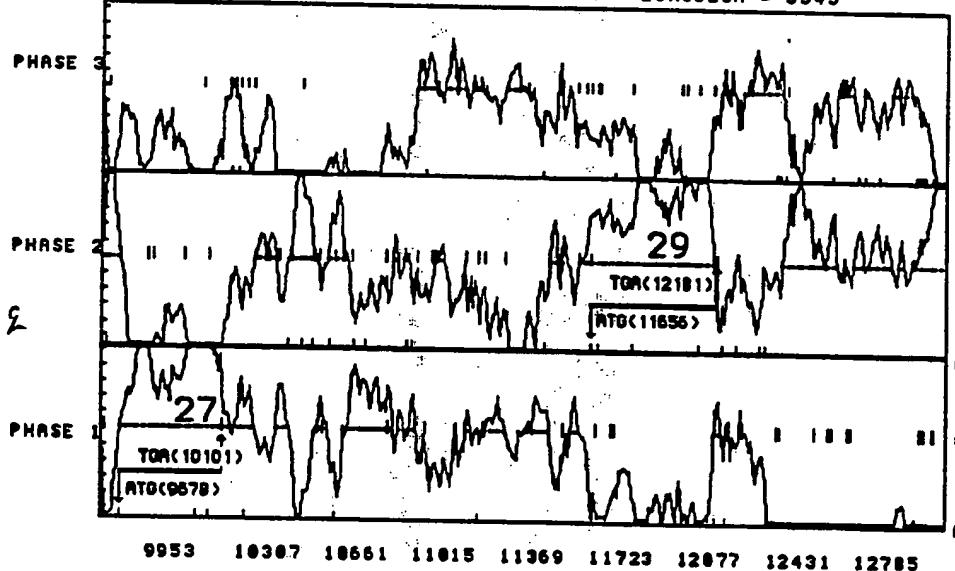
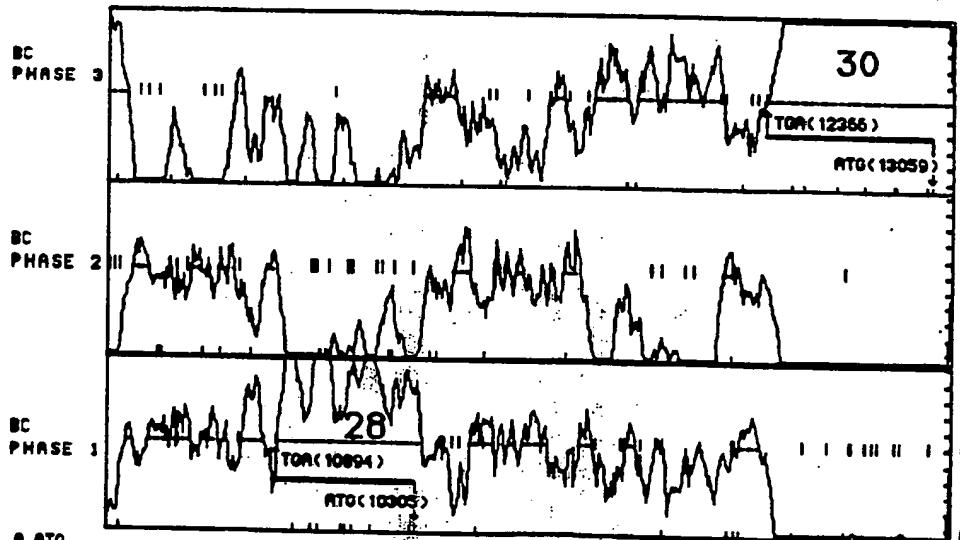
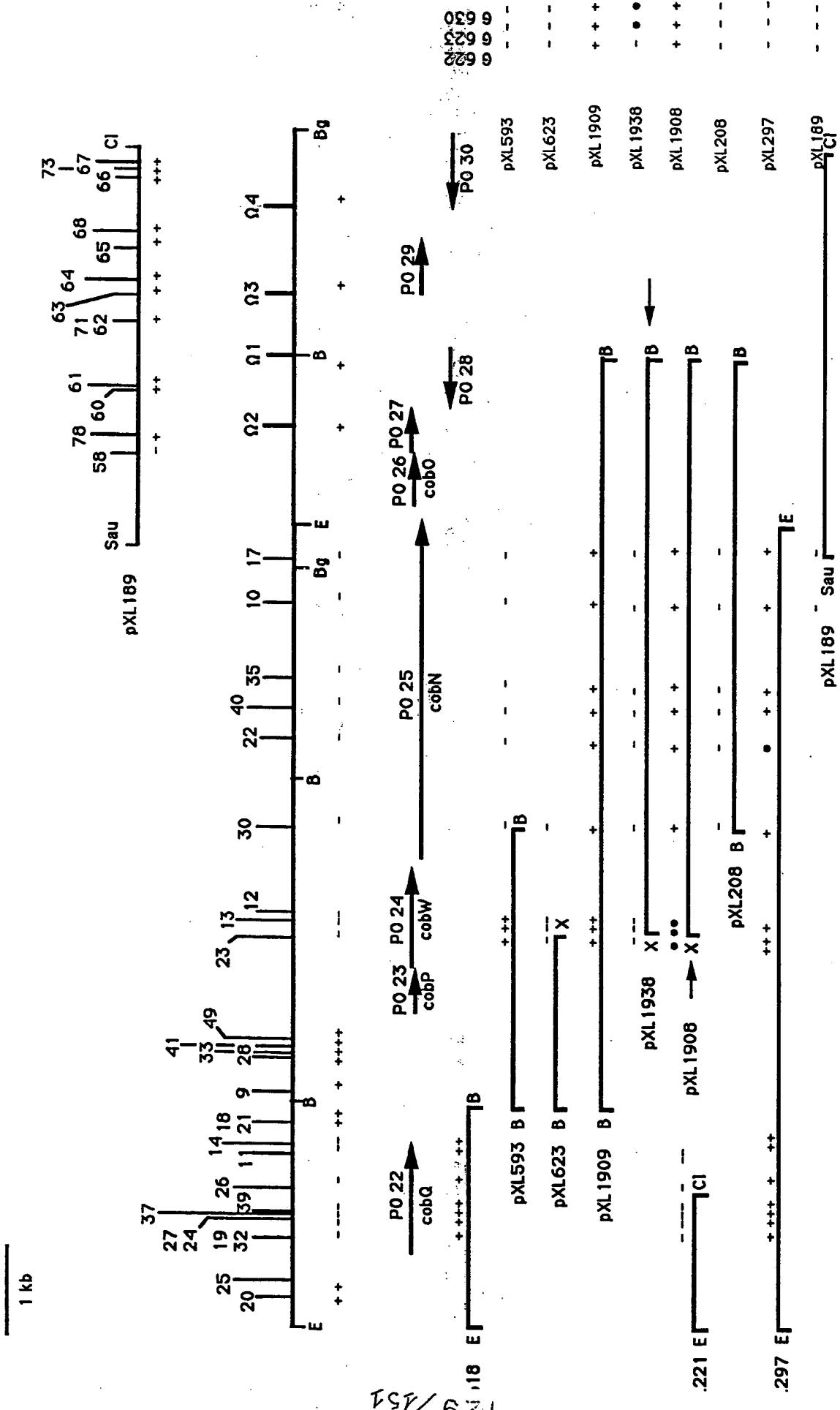


FIG. 45.2



PHASES OUVERTES 26, 27, 28, 29, 30

FIGURE 46



PCT/FR91/00054

W0911/1518

130

1451

Sequence de 13144 pb de 429 a 1886 gène cobQ

M T R R I M L Q G T G S D U G K S U L U A G L C R  
ATGACACGCAGGATCATGGTCAGGGAACTGGCTCGGATGTCGGAAATCGGTATTGGTGGCGGGCTCTGCCGG  
429 439 449 459 469 479 489 499  
L A A N Q G L K U R P F K P Q H H S H N A A U S D  
CTTGCCGCCAATCAGGGCCTGAAAGTCCGCCGTTCAAGCGGCAGAACATGTCGAACACGCGCCGCGTTCCGAC  
504 514 524 534 544 554 564 574  
D G G E I G R A Q W L Q A L A A R U P S S U H M H  
GACGGCGGGCAGATCGGCCGCGCAGATGCGCTGCAGGGCGCTGGCCGCGCGTCCGTCGGTGCACATGRC  
579 589 599 609 619 629 639 649  
P U L L K P Q S D U G S O I U U Q G K U A G Q A R  
CCGGTGCTCCTGAAAGCCGCACTGGCGCTGGACGCTGGCGCCGTCATGGAGAGTTTCGACACAAATATCGGCCG  
654 664 674 684 694 704 714 724  
G R E Y Q A L K P K L L G A U M E S F E O I S A G  
GGGCAGGAAATATCAGGCCTCAAGCCCAGCTGCTGGCGCCGTCATGGAGAGTTTCGACACAAATATCGGCCG  
729 739 749 759 769 779 789 799  
A D L U U U E G R A G S P A E I M L R P G D I A H M  
GCCGATCTCGTGGTGGTCGAGGGCGCCGCTGCCGCCGAAATCACCTCAAGGCCCGGACATGCCAAATATG  
804 814 824 834 844 854 864 874  
G F A T R A H U P U L U G D I D R G G U I A S L  
GGCTTGCGACACGGGCCAATGTCGCEGGTCTGCTGGCGACATCGACCGGGGGGGGTGATCGCTCGCTG  
879 889 899 909 919 929 939 949  
U G T H A I L P E E D R R M U T G Y L I N K F R G  
GTCGGCACGCATGCATCCTGCCGAGGGAGACCGGGCCTGGTGCACGGCTATCTCATCACACAGTTCCGCGGC  
954 964 974 984 994 1004 1014 1024  
D U T L F D D G I A A U H R Y T G H P C F G U U P  
GRCGTACGCTGTTGACGACGGCATGGCTGCCGCTACCGGCTACCCGGCTGGCCCTGCTTCGGCGTGTGCG  
1029 1039 1049 1059 1069 1079 1089 1099  
H L K A A A R L P A E D S U U L E K L T R G E G R  
TGGCTGAGGGCGGGCGACGCCGCTGCCGGGAGGATTCGCTGCTGGAGAGCTGACGCCGGCGAGGGCG  
1104 1114 1124 1134 1144 1154 1164 1174  
A L K U A U P U L S R I A N F D D L D P L A A E P  
GCGCTGAGGGTTGCCGCTCCGGTACTGTCGCATGCCAATTCGACGCCATCCGATCCGCTGCCGCCGACCG  
1179 1189 1199 1209 1219 1229 1239 1249  
E I D L U F U R P G S P I P U D A G L U U I P G S  
GAGATTGATCTCGTCTCGTGGCGCTGGCAGTCCTCCGGTGCACGCTGGCCTCGTCGTCATTCCGGCTG  
1254 1264 1274 1284 1294 1304 1314 1324  
K S T I G D L I D F R A Q G H D R D L E R H U R R  
AAATCGACCATCGGCCACCTCATCGATTTCCGTCGCAAGGGTGGGACCGTGCACCTCGAACGTCATGTGCGCCGG  
1329 1339 1349 1359 1369 1379 1389 1399  
G G R U I G I C G G V Q H L G R R U T D P L G I E  
GGCGGCCGGGTCACTGGCATCTGCGGGCTACCGAGATGCTGGCCGGCGCTACCGATCCGATCCGCTCGGCATCGAG  
1404 1414 1424 1434 1444 1454 1464 1474  
G G E R A U E G L G L L E U E T E M A P E K T U R  
GGCGGCCGACGTGCGGTGAGGGCCTCGGGCTGCTCGAGGTCGAGACCGAGATGGCGCCGGAAAAGACGGTGC  
1479 1489 1499 1509 1519 1529 1539 1549  
H S A U W S L E H D U U L E G Y E I H L G K T Q G  
AACAGCCGCGCCTGGTCGCTGGAGCATGATGTGGTGCACGGCTACGAAATCCATCTGGCAAGACGCAAGGT  
1554 1564 1574 1584 1594 1604 1614 1624  
A D C G R P S U R I D H R A D G A L S A D G R U M  
GCGGACTGTGGCCGGCGTCGGTGCACATCGAACATCGGCCGCGCCCTTCGGCCGATGGCCGCGTGTG  
1629 1639 1649 1659 1669 1679 1689 1699  
G T Y L H G L F T S D A Y R G A L L K S F G I E G  
GGCACCTACCTGCATGGCTCTCACCGAGCGCCATCGCGGCCGCGCTGCTCAGAGAGTTTCGGCATCGAAGGC  
1704 1714 1724 1734 1744 1754 1764 1774  
G A H M Y R Q S U D A A L D D U A N E L E A U L D  
GGCGCCACACACTACCGCCACATCGGTGATGCGCGCTCGACGATGTCGCAACGARCTGGAGGCTGTGCTCG  
1779 1789 1799 1809 1819 1829 1839 1849  
R R W L D E L L R H \*  
CGTCGCTGGCTGGACGAGTTGCTCAGGCACTAG  
1854 1864 1874 1884

131 / 151

## PROTEINE COBQ

PREMIER RESIDU = 1  
DERNIER RESIDU = 485

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	11	2.27	1617.75	3.11
2	LEU	L	50	10.31	5654.20	10.88
3	ILE	I	23	4.74	2600.93	5.01
4	MET	M	10	2.06	1310.41	2.52
5	VAL	V	50	10.31	4953.42	9.53
6	SER	S	24	4.95	2088.77	4.02
7	PRO	P	23	4.74	2232.21	4.30
8	THR	T	15	3.09	1515.72	2.92
9	ALA	A	49	10.10	3480.82	6.70
10	TYR	Y	8	1.65	1304.51	2.51
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	1.44	959.41	1.85
13	GLN	Q	15	3.09	1920.88	3.70
14	ASN	N	16	3.30	1824.69	3.51
15	LYS	K	15	3.09	1921.42	3.70
16	ASP	D	34	7.01	3910.92	7.53
17	GLU	E	28	5.77	3613.19	6.96
18	CYS	C	4	0.82	412.04	0.79
19	TRP	W	6	1.24	1116.48	2.15
20	ARG	R	40	8.25	6244.04	12.02
21	GLY	G	57	11.75	3250.22	6.26
22	-	-	0	0.00	0.00	0.00

RESIDUS = 485

Masse moleculaire (monoisotopique) = 51950.1016

Masse moleculaire (moyenne) = 51982.3711

INDEX DE POLARITE (%) = 40.00

POINT ISOELECTRIQUE = 6.16

DD 260 (1mg/ml) = 0.558 DD 280 (1mg/ml) = 0.825

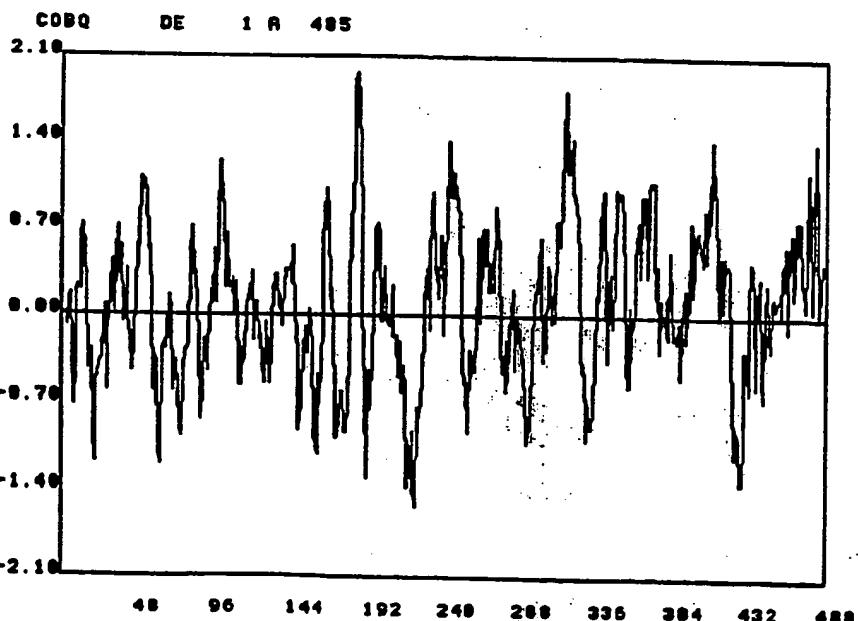


FIGURE 47.1

132/151

Sequence de 13144 pb de 3364 à 3888 gène cobP

M S S L S A G P U L U L G G A R R S G K S S F S E R							
ATGAGCGAGTCTCAGCGCCGGGCCGTGCTGGTCCTGGCGGCCGTTCCGGCAGTCCAGCTTTCCGAGAGG							
3364 3371 3384 3394 3404 3414 3424 3434							
L U E A S G F T M H Y U A T G R A W D D E M R E R							
CTCGTCCGAGCGTCCGGCTTCACCATGCATTATGTCGCCACGGGCCGCTGGGACGACGAAATGCGCGAGCGC							
3439 3449 3459 3469 3479 3489 3499 3509							
I D H H R T R R G E G U T T H E E P L D L U G I L							
ATCGACCCATCACCGGACCGCGCCGGCGAGGGCTGGACGACGATGAGGGAGCCGCTCGATCTCGTCGGCATCCTC							
3514 3524 3534 3544 3554 3564 3574 3584							
R R I D D P S H U U L I D C L T L U U T N L M L E							
AGACGCGCATCGATGATCCCAGCCATGTGGTCTGTGACTGCCTGACCGCTATGGGTCAACCATCTCATGCTGGAA							
3589 3599 3609 3619 3629 3639 3649 3659							
E R D M T A E F R A A L U A Y L P E A R A A R L U F U							
GRGCGCGACATGACGGCGGAGTTGCGCCGCTTGTTGCGTATCTGCCCGAGGGCGCGGGCGCCTCGTCTTGT							
3664 3674 3684 3694 3704 3714 3724 3734							
S N E U G L G I U P E H R M A R E F R D H A G R L							
TCCCATGAGGTGGCCCTCGGCATCGTGCCTGAGAACCGCATGGCCCGAGTTGCGACCCATGCCGGCCGGCTT							
3739 3749 3759 3769 3779 3789 3799 3809							
H Q I U A E K S R E U Y F U A A G L P L K M K G *							
CACCAAGATCGTTGCGGAGAAATCCGCTGAAAGTTACTTGTGCGGGCCGGTTGCGCTGAAATGARGGGTTGA							
3814 3824 3834 3844 3854 3864 3874 3884							

FIGURE 47.2

133/151

## PROTEINE COBP

PREMIER RESIDU = 1  
DERNIER RESIDU = 174

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	6	3.45	882.41	4.54
2	LEU	L	19	10.92	2148.60	11.06
3	ILE	I	6	3.45	678.50	3.49
4	MET	M	7	4.02	917.28	4.72
5	VAL	V	16	9.20	1585.09	8.16
6	SER	S	11	6.32	957.35	4.93
7	PRO	P	6	3.45	582.32	3.00
8	THR	T	8	4.60	808.38	4.16
9	ALA	A	17	9.77	1207.63	6.22
10	TYR	Y	3	1.72	489.19	2.52
11	*	*	0	0.00	0.00	0.00
12	HIS	H	7	4.02	959.41	4.94
13	GLN	Q	1	0.57	128.06	0.66
14	ASN	N	3	1.72	342.13	1.76
15	LYS	K	4	2.30	512.38	2.64
16	ASP	D	9	5.17	1035.24	5.33
17	GLU	E	16	9.20	2064.68	10.63
18	CYS	C	1	0.57	103.01	0.53
19	TRP	W	3	1.72	558.24	2.87
20	ARG	R	17	9.77	2653.72	13.66
21	GLY	G	14	8.05	798.30	4.11
22	-	-	0	0.00	0.00	0.00

RESIDUS = 174

Masse moleculaire (monoisotopique) = 19429.9473

Masse moleculaire (moyenne) = 19442.2637

INDEX DE POLARITE (%) = 43.68

POINT ISOELECTRIQUE = 6.71

DD 260 (1mg/ml) = 0.720 DD 280 (1mg/ml) = 1.042

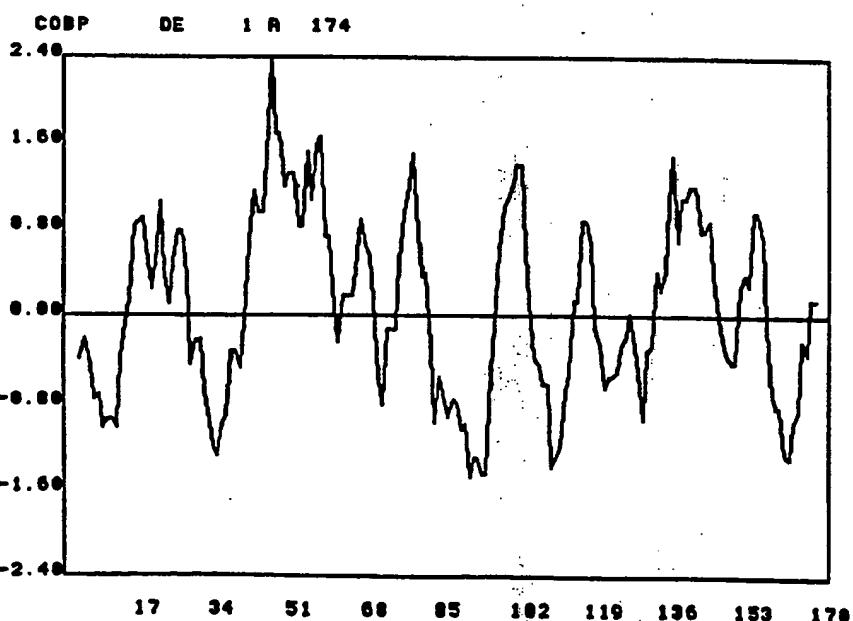


FIGURE 47.3

134 / 151

Sequence de 13144 pb de 3892 à 4956 gène cobH

M T T A R A M Q G K I P A T U I T G F L G A G K T  
 ATGACCACTGCGAGGCCAACCGAGGGCAAGATCCGGCAGCCGTATCACCGGCTTCTCGGCAGCCGGCAAGACG  
 3892 3902 3912 3922 3932 3942 3952 3962  
 T M I R N L L Q H A D G K B I G L I I N E F G D L  
 ACGATGATCCGCAACCTGCTGCAGAACGCCAGCGCARGCGCATCGGCCTGATCATCACGAGTTGGCGATCTT  
 3967 3977 3987 3997 4007 4017 4027 4037  
 G U D G D U L K G C G A E A C T E D O D I I E L T H  
 GGCCTGCGACGGCGRTGCTTGRAGGGCTGGCTGCCGAGGGCTGCACCGAGGGACGACATCATCGAGCTCACCRAT  
 4042 4052 4062 4072 4082 4092 4102 4112  
 G C I C C T U A D D F I P T M T K L L E R E N R P  
 GGCTGCATCTGCTGCACCGTGGCTGACGATTCATCCGACCATGACGAGCTGCTCGAGCGTGAACACCGCTCCT  
 4117 4127 4137 4147 4157 4167 4177 4187  
 D H I I I E T S G L A L P O P L I A A F H W P D I  
 GACCCACATCATCATCGAACCTCGGGCCTTGCCCTGCCGAGCCGCTGATCGCCGCTTCRACGGCCGGATATC  
 4192 4202 4212 4222 4232 4242 4252 4262  
 R S E U T U D G U U T U U D S A A U A A G R F A D  
 CGCAGCGAAGTGAACCGTCGATGGCGTCGTACCGTGGTCGACGEGCCGCCGTTGCCGCTGGCCGCTTGCCGAC  
 4267 4277 4287 4297 4307 4317 4327 4337  
 D H D K U D A L R U E D D H L D H E S P I E E L F  
 GACCCACGACRGGTGATGCGCTGCCGTCGAGGACGACATCTGATCACGAAAGCCCCGATCGAGGAGCTGTT  
 4342 4352 4362 4372 4382 4392 4402 4412  
 E D Q L T A A D L I U L H K T D L I D A S G L K A  
 GRGGATCAACTGACGGCTGCCGATCTCATCGTTCTCAGAACAGCCGATCTGATGCGATGCCCTCCGGCCTCAGGCC  
 4417 4427 4437 4447 4457 4467 4477 4487  
 U R D E U S S R T S R K P T M I E A K H G E U A A  
 GTGCGCGACGAGGTGTCTCGCGCACAGCCGACGAGGCCACGATGATCGAGGCGAAACCGGCAAGTGCACCGCT  
 4492 4502 4512 4522 4532 4542 4552 4562  
 A I L L G L G U G T E S D I A H R K S H H E M E H  
 GCCATCCTGCTGGCCTCGGTGTGGCACGGAAAGCGATATGCCAACCGAACGATCGCCGATCGACGAGATGGAGCAC  
 4567 4577 4587 4597 4607 4617 4627 4637  
 E A G E E H D H D E F D S F U U E L G S I A D P A  
 GAGGGCAGGTGAGGAGCAGATCACGACGAGATCGAGCTCGCTGAGCTCGGGTCGATCGCCGATCCGGC  
 4642 4652 4662 4672 4682 4692 4702 4712  
 A F I D R A L K G U I A E H D U L R L K G F A D U P  
 GCCTTCATCGATGCCCTGAGGGCGTAATCGCGGAGCACGAGCTCGCTGAGCTCGGGTCGATCGCCGATCCGGC  
 4717 4727 4737 4747 4757 4767 4777 4787  
 G K P M R L L I Q A U G A R I D Q Y Y D R A W G A  
 GGCAAGCCGATGCGCCTCTGATCCAGGCGGTGGCGCCGATCGACCAATTACGACGCCCTGGGGCTGGGGCGCT  
 4792 4802 4812 4822 4832 4842 4852 4862  
 G E K R G T R L U U I G L H D M D E A A U R A A I  
 GGCAGAAAGCGCGGTACGCGCCTCGTCGTATCGGCCTGCACGACATGGACGAGGGCGGGTGCAGCGCCGATC  
 4867 4877 4887 4897 4907 4917 4927 4937  
 T A L U \*  
 ACCCGCCTCGTAG

FIGURE 47.4

135 / 351

## PROTEINE COBH

PREMIER RESIDU = 1  
DERNIER RESIDU = 354

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	10	2.82	1470.68	3.86
2	LEU	L	32	9.04	3618.69	9.50
3	ILE	I	28	7.91	3166.35	8.31
4	MET	M	7	1.98	917.28	2.41
5	VAL	V	28	7.91	2773.92	7.28
6	SER	S	12	3.39	1044.38	2.74
7	PRO	P	11	3.11	1067.58	2.80
8	THR	T	21	5.93	2122.00	5.57
9	ALA	A	41	11.58	2912.52	7.64
10	TYR	Y	2	0.56	326.13	0.86
11	*	*	0	0.00	0.00	0.00
12	HIS	H	10	2.82	1370.59	3.60
13	GLN	Q	6	1.69	768.35	2.02
14	ASN	N	11	3.11	1254.47	3.29
15	LYS	K	15	4.24	1921.42	5.04
16	RSP	D	36	10.17	4140.97	10.87
17	GLU	E	27	7.63	3484.15	9.15
18	CYS	C	5	1.41	515.05	1.35
19	TRP	W	2	0.56	372.16	0.98
20	ARG	R	20	5.65	3122.02	8.19
21	GLY	G	30	8.47	1710.64	4.49
22	-	-	0	0.00	0.00	0.00

## RESIDUS = 354

Masses moléculaires (monoisotopique) = 38097.4258  
 Masses moléculaires (moyenne) = 38121.1055  
 INDEX DE POLARITÉ (%) = 44.63  
 POINT ISOELECTRIQUE = 4.90  
 D<sub>O</sub> 260 (1mg/ml) = 0.268 D<sub>O</sub> 280 (1mg/ml) = 0.354

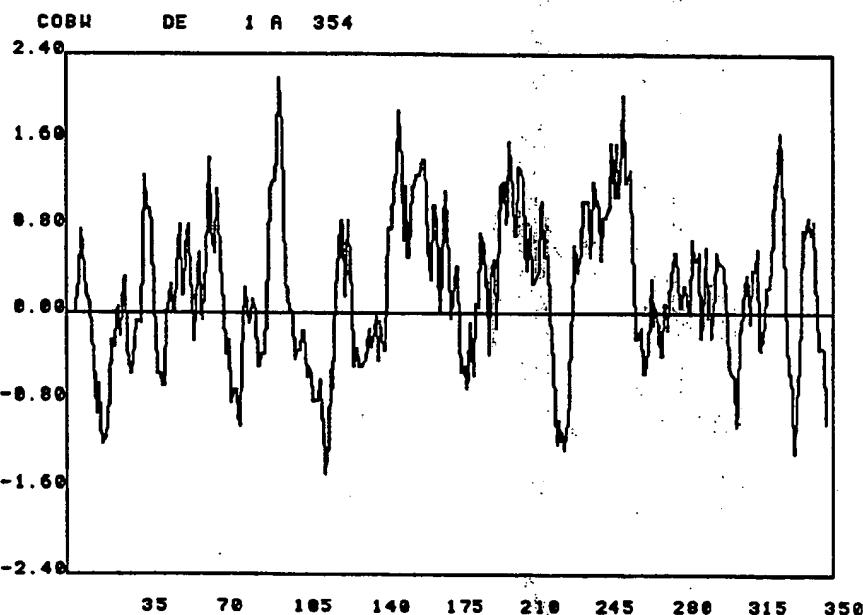


FIGURE 47.5

136/151

Sequence de 13144 pb de 5060 à 6887 gène cobM

M H L L L A Q K G T I A D G N E A I D L G Q T P A  
ATGCATCTGCTTCTGCCCAAGAARRGGARCGATCGCCGACGGCRACGAGGCATCGACCTTGGGCAAAACGCCGGCC  
5060 5070 5080 5090 5100 5110 5120 5130  
D I L F L S R A D T E L S S I A A A A H G R R D G G  
GATATCCTTTCTATCGGCCGCCGACACCGAGCTCTCCTCGATCCCGCAGGCTCACGGCCGACCGACCGAGGC  
5135 5145 5155 5165 5175 5185 5195 5205  
L S L R I A S L M S L M H P M S U D T Y U E R T A  
TTGAGCCTGCGCATCGCCAGCCTGATGAGCCTGATGCAACCCGATGTCGGTCGACACTTACGTGAGCGCACGGCG  
5210 5220 5230 5240 5250 5260 5270 5280  
R H A K L I U U R P E G G A S Y F R Y L L E A L H  
CGTCACGCCAAGCTGATCGTCGTCGGCCGCTGGTGGCGCAGCTATTCCGTTCTGCTGGRAAGCCCTGCAT  
5285 5295 5305 5315 5325 5335 5345 5355  
A A A U T H R F E I A U L P G D D K P D P G L E P  
GCGGCTGCCGTACCCATGTTTCAGAGATCGCGGTTCTGCCGGGTGACGACAGCCGGATCCGGGCTGGAGGCC  
5360 5370 5380 5390 5400 5410 5420 5430  
F S T U A A D D R Q A L H A Y F T E G G S D H A G  
TTCTCCACCGTCGCGCCGACGACCGCCAGCGCCCTTGGGCTACTCACCGAAGGCGCTCGGACAAATGCCGG  
5435 5445 5455 5465 5475 5485 5495 5505  
L F L D Y A A A L U T G A E K P Q P A K P L L K A  
CTGTTCTCGACTATCGGGCCGRCCTGGTCAGGTGCGGAGAGCCGAGCCCCGAAAGCCCCTGTTGAGGCC  
5510 5520 5530 5540 5550 5560 5570 5580  
G I H W P G A G U I G U S E H Q S L U Q G R M U A  
GGCATCTGGTGGCCGGGTGCTGGTGTATCGCGCTCACCGAATGGCAGTCCTTGTCAGGGACGGATGGTAGCG  
5585 5595 5605 5615 5625 5635 5645 5655  
R E G F E P P T U G I C F Y R A L U Q S G E T R P  
AGGGAGGGATTGARACCCCCGACGGTCGGGATCTGCTTTACCGCGCTCGTCAGAGTGGCGAGACACGGCC  
5660 5670 5680 5690 5700 5710 5720 5730  
U E A L I D A L E A E G U R A L P U F U S S L K D  
GTGGAGGCCTGATCGCTGGAGGCTGARGGTGTGCGGGCAGCTGCCGGTGTCTCAAGCCTCAAGGAT  
5735 5745 5755 5765 5775 5785 5795 5805  
A U S U G T L Q A I P S E A A P D U U M M A T G F  
GCCGTTCCGTCGGCACGCTGCAGGGCATTTCCGAGGGCCGACGTGGTGTGAAACGCCACTGGCTT  
5810 5820 5830 5840 5850 5860 5870 5880  
A U S S P G A D R Q P T U L E S T G A P U L Q U I  
GCCGTTCTCGTCGCCGGTGCCGACCGTCAGCCGACGGTGTGGAAATCGACCCGGTGCAGGGTGTGATT  
5885 5895 5905 5915 5925 5935 5945 5955  
F S G S S R A Q H E T S P Q G L M A R D L A M H U  
TTCTCCGGCTCGCGGGCGCAATGGGAAACGTGCGCCGAGGGCTGATGGCGCGACCTCGCCATGAAACGTG  
5960 5970 5980 5990 6000 6010 6020 6030  
A L P E U D G R I L S R A U S F K A A S I Y D A K  
GCACCTCCCCGAGTCGATGGCCGATCCTTCGCGCGCCGTCCTCCTCAAGGGCGCTGATCTATGACGCCAG  
6035 6045 6055 6065 6075 6085 6095 6105  
U E A M I V G H E P L E G R U R F A A D L A U M U  
GTGGAGGCCAATATCGCGCCATGAGCCGCTCGAAGGGCGGGTGCGCTTGCCTGATCTTGCCGTCACTGG  
6110 6120 6130 6140 6150 6160 6170 6180  
A N U R R A E P A E R R I A I U M A H Y P N R D G  
GCGAACGTGCGCCGGGAGAGCCGCCGAGCGCCGTATGCCATCGTCATGGCAGACTATCCGAAACCGCGACGG  
6185 6195 6205 6215 6225 6235 6245 6255  
R L G H G U G L D T P A G T U E U L S A M A R E G  
CGCCTCGGCAACGGTGTGGGCTCGACACGCCGGCGTACCGTCGAGGTGCTTAGCGCCATGGCGGGAGGC  
6260 6270 6280 6290 6300 6310 6320 6330  
Y A U G E U P A D G D A L I R F L M A G P T H A A  
TATGCGGTGCGGTGAGGGTCCCGCCGATGGCAGCGCGCTGATCCGCTTCTGATGGCGGGCGACCAATGCGGCG  
6335 6345 6355 6365 6375 6385 6395 6405  
S H D R E I R E R I S L N D Y K T F F D S L P K Q  
AGCCATGACCGTGAATCCCGCGAGCGCTGATTTGCTGACCGATTACAAAACGTTCTCGATTCGCTTCCGAAACAG  
6410 6420 6430 6440 6450 6460 6470 6480  
I K D E U A G R H G U P E A D P F F L D G A F A L  
ATAAGGATGAAAGTTGCCGGTGCCTGGGGCGTGCAGGGCCGATCCCTTTCTCGATGGCGCCTTCGCGCTG  
6485 6495 6505 6515 6525 6535 6545 6555

FIGURE 47.6

137/151

P L A R F G E U I U G I Q P A R G Y H I D P K E S  
 CCGCTCGCCCGCTTCGGCGAGGTGATCGTCGGCATCCAAACCGGCGCGCGCTACRACATCGATCCGAAGGGARAGC  
 6560 6570 6580 6590 6600 6610 6620 6630  
 Y H S P D L U P P H G Y L A F Y A F L R Q Q F G A  
 TACCATCCCCGGACCTCGTCCGCCGCATGGCTATCTGCCCTCTACGCCCTCGGCCAGCAGTCGGAGCG  
 6635 6645 6655 6665 6675 6685 6695 6705  
 Q A I U H M G K H G N L E W L P G K A L A L S E T  
 CAGGCGATCGTCCACATGGGCAGCGCACGGCATCTCGAATGGCTGCCGGCAGGGCGCTGGCGCTGCGGAAAC  
 6710 6720 6730 6740 6750 6760 6770 6780  
 C Y P E A I F G P L P H I Y P F I U N D P G E G T  
 TGCTATCCCAGCGATCTTCGGGCCGCTGCCGCACATCTATCCCTCATCGTCAACGATCCGGCGAGGGTACG  
 6785 6795 6805 6815 6825 6835 6845 6855  
 Q A K R R T S A U I L D H L T P P L T R A E S Y G  
 CAGGCCAGCGCCGCCACRGCGCCGTCATCGACCCACCTGACCCCGCCCTTGACCGCGCCGAAATCCATCGGC  
 6860 6870 6880 6890 6900 6910 6920 6930  
 P L K D L E A L U D E Y Y D A R A G G D P R R L R L  
 CCGCTCAAGGATCTGGAGCGCTCGTCGACGAAATTACGAGCCGGCGGTGATCCGCGCCGCTCAGGCTG  
 6935 6945 6955 6965 6975 6985 6995 7005  
 L S R Q I L D L U R D I G L D S D A G I D R G D S  
 CTCAGCCGCCAGATCCTCGATCTCGTCGCGACATCGGCCCTCGACAGCGACGGCATCGACAGGGCGACAGC  
 7010 7020 7030 7040 7050 7060 7070 7080  
 D D K A L E K L D A V L C D L K E M Q I R D G L H  
 GACGACAGGGCGCTGGAGGAGCTCGACGCCATCTCGCGACCTCAAGGAAATGCAAGATCCGCGACGGCCTGCAC  
 7085 7095 7105 7115 7125 7135 7145 7155  
 I F G U A P E G R L L T D L T U A L A R U P R G L  
 ATCTTCGGCGTTGCGCCGGRAAGGGCGGTTGTCAGGGACCTCACCGTAGCGCTGGCGCGTGCCTGGGGAGGGTCTC  
 7160 7170 7180 7190 7200 7210 7220 7230  
 G E G G D Q S L Q R A I A A D A G L R G F A I P T  
 GGCAGAGGGCGGCCAGAGCCTCGACGGCGATCGCAGCGGATGCGGGGCTCGTGGTTGCTATTCCACC  
 7235 7245 7255 7265 7275 7285 7295 7305  
 S A G G H P A R D A Q P F D P L D C U M S D T H T  
 TCGGCAGGGGGCAACCCCGCACGCCAACCCCTTCGACCCGCTCGACTCGCTCATGTCGACACCTGGAC  
 7310 7320 7330 7340 7350 7360 7370 7380  
 G P K P S I L A D L S D A P H R T A G D T U E R I  
 GGCCCCGAAACCGTCCATCCTCGCTGACCTCTCGGACGCCCTGGCGACCGCCGGGATACGGTCGAGCGCATC  
 7385 7395 7405 7415 7425 7435 7445 7455  
 E L L A A H L U S G E L A C P O H H A N T R A U L  
 GAGTTCTGCGCAATCTCGTGTGGGTGAACTGGCTTGCCTGGGCAACACCCGGCGCCGTGCTC  
 7460 7470 7480 7490 7500 7510 7520 7530  
 G E I E T R L K P S I S M S G A A E M T G F L T G  
 GGCAGAAATCGAAGCGCCCTGAGGGCGTCGATTTCAACACTGGGTGCCCGAGATGACCGGCTTCCTCACCGGT  
 7535 7545 7555 7565 7575 7585 7595 7605  
 L S G R F U A P G P S G A P T R G R P D U L P T G  
 CTCAGCCGCCGCTCGCGCCCCGGTCATCGGCCGCGCCGACGCCGGGATGTGTTGCCGACGGGG  
 7610 7620 7630 7640 7650 7660 7670 7680  
 R H F Y S U D S R A U P T P A A Y E L G K K S R E  
 CGCAATTCTACTCGGTGACAGCCGEGCGTGCACGCCGGCTACGAGCTGGCAGAAATCGGCCGAG  
 7685 7695 7705 7715 7725 7735 7745 7755  
 L L I R R R Y L Q D H G E W P S S F G L T A H G T A  
 CTTCTGATCCGCCGCTACCTCGAGGACCATGGCGATGGCCGTCTCCCTGGCCTGACCGCCTGGGGCACGGCG  
 7760 7770 7780 7790 7800 7810 7820 7830  
 H M R T G G D I A Q A L R L I G A K P T H D M U  
 AACATCGCGCACCGGGCGACGACATCGCCAGGGCCCTGGCGCTGATCGGCCAGCCACCTGGGACATGGTC  
 7835 7845 7855 7865 7875 7885 7895 7905  
 S R R U M G Y E I U P L A U L G R P R U D U T L R  
 TCTCGCCGGGTGATGGGCTACGAGATCGTGGCGCTCGCAGTCCTCGCCGCCACCGCGTCGACGTGACCTTCGCG  
 7910 7920 7930 7940 7950 7960 7970 7980  
 I S G F F R D A F P D O I A L F D K A I R A U A L  
 ATTCGGCTTCTCCGCGATGCCCTCCGGACCGAGATCGCGCTCTCGACAGGGCGATCCGCGCCGTGCGCGT  
 7985 7995 8005 8015 8025 8035 8045 8055  
 E E D D A D H M M I A A R M A R A E S R R L E A E G U  
 GGGGAGCGATGCCGACACATGATGCCGACCGCATGCCGGCGGAGCCGGCTGGAGGGCGAGGGCGT  
 8060 8070 8080 8090 8100 8110 8120 8130

138 / 151

E A A E A A R R A S Y R U F G A K P G A Y G A R A L  
 6AAGCCGCCGAGGCCGCGCGTCGGCGCTCCATACCGCGTCTTGCGCARRAGCCCGGTGCCTATGGCGCCGCCCTG  
 8135 8145 8155 8165 8175 8185 8195 8205  
 Q A L I D E K G H E T K A D L A E A Y L T H G A Y  
 CAGGCCTGATCGRCGAGAAGGGCTGGGAAACCGAACAGATCTCGCCGAGGCCTATCTTACCTGGGGCGCCCTAT  
 8210 8220 8230 8240 8250 8260 8270 8280  
 A Y G A G E E G K A E R D L F E E R L R T I E A U  
 GCCTATGGCGCCGGCGRGGAGGGCAAGGCCGAGCGATCTTCGAGGAGCGCCTGCGCACGATAGAGGCCTG  
 8285 8295 8305 8315 8325 8335 8345 8355  
 V Q H Q D H R E H D L L D S D D Y Q F E G G M S  
 GTGCAGAACCGGACRCCCGCGAGCACGATCTGCTCGACAGCAGACTACTACCCAGTTGRRAGGCCTGATGAGC  
 8360 8370 8380 8390 8400 8410 8420 8430  
 R A A E Q L G G H R P R I Y H N D H S R P E K P U  
 GCTGCCGCCGAAACGCTCGGCCTCACCGTCCGGCGATCTACCCACACGACCATTCCCGTCCGGAAAAAGCCTGTG  
 8435 8445 8455 8465 8475 8485 8495 8505  
 I R S L E E E I G R U U R A R U V U N P K H I D G U  
 ATCCGGTCGCTCGAGAGAAGAGATCGGCCGGTGGTCCGGGCCCCTCGTCAATCCCGAGTGGATCGATGGCGTC  
 8510 8520 8530 8540 8550 8560 8570 8580  
 M R H G Y K G A F E I R A T U D Y M F A F A A T T  
 ATGCGCCACGGATACAGGGCGCCCTCGAGATCGCTGCCACGGTCGACTACATGTTGCCCTTGCCGCAACCG  
 8585 8595 8605 8615 8625 8635 8645 8655  
 G A U R D H H F E A A Y Q A F I U D E R U A D F M  
 GGTGCGGGTGCACGACCATCTTCGAGGCCGCTTATCAGGGCGTTCTTCGACGAGCGCTGGCTGACTTCATG  
 8660 8670 8680 8690 8700 8710 8720 8730  
 R D K H P A A F A E L A E A L L E A I D R H L W T  
 CGCGACAAAGAACCCGGCCCTTGCCGAGCTTGCCGAACGCCCTGCTGAGCCTCGACCCCAATCTCTGGACG  
 8735 8745 8755 8765 8775 8785 8795 8805  
 P R S H S A A F E L R G I G T A A T R L R A G H E  
 CGCGCTCGAATTGGCGCGGTTGAACTTGCCTGGCACGGCAGCAGAACCCGGCTTGTGCCGGCAATGAA  
 8810 8820 8830 8840 8850 8860 8870 8880

\*  
TAG  
8885

FIGURE 47.8

139/151

PROTEINE COBN			PREMIER RESIDU =	1
			DERNIER RESIDU =	1275
		NOMBRE	X NOMB	POIDS
1	PHE	F	48	3.76
2	LEU	L	121	9.49
3	ILE	I	60	4.71
4	MET	M	24	1.88
5	VAL	V	82	6.43
6	SER	S	64	5.02
7	PRO	P	76	5.96
8	THR	T	53	4.16
9	ALA	A	180	14.12
10	TYR	Y	35	2.75
11	*	*	0	0.00
12	HIS	H	24	1.88
13	GLN	Q	32	2.51
14	ASN	N	30	2.35
15	LYS	K	34	2.67
16	ASP	D	90	7.06
17	GLU	E	85	6.67
18	CYS	C	5	0.39
19	TRP	W	18	1.41
20	ARG	R	99	7.76
21	GLY	G	115	9.02
22	-	-	0	0.00
				0.00

RESIDUS = 1275  
 Masse moleculaire (monoisotopique) = 137970.5000  
 Masse moleculaire (moyenne) = 138055.8594  
 INDEX DE POLARITE (%) = 40.08  
 POINT ISOELECTRIQUE = 5.42  
 DO 260 (mg/ml) = 0.693 DO 280 (mg/ml) = 1.027

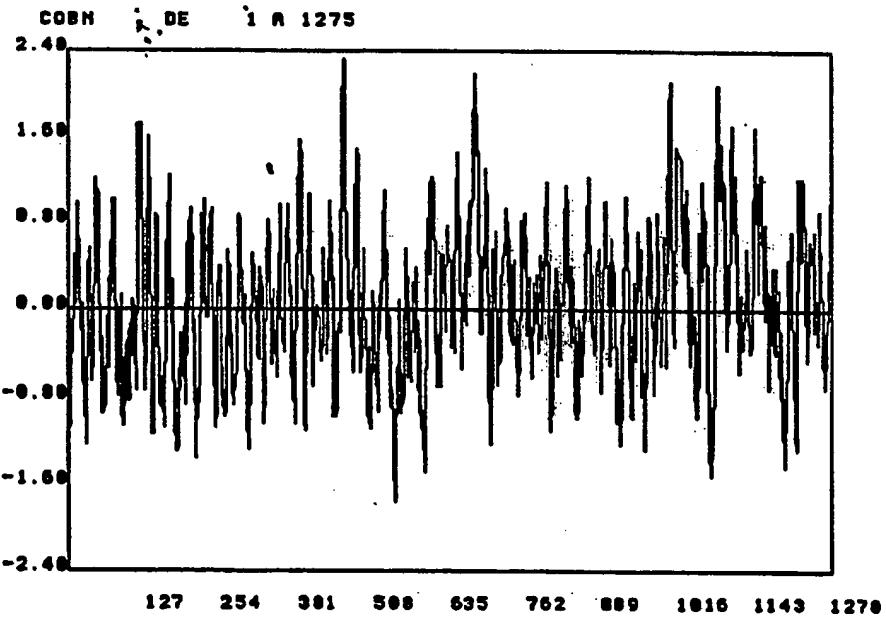


FIGURE 47.9

140/158

Sequence de 13144 pb    9034 à    9678    gène cobO

M S D E T T U G G E A P A E K D D R R H A M K M A  
ATGAGCGACGAGACGACAGTAGGCAGGCGAGGCCGGAGAAGGACGATGCCCGCCACGCCATGAGATGGCG  
9034 9044 9054 9064 9074 9084 9094 9104  
K K K A A R E K I M A T K T D E K G L I I U N T G  
AAGAAGAGGGCAGCACGCGAAGGATCATGGCGACGAGAGGAGGGTCTGATCATCGTCACACCCGGC  
9109 9119 9129 9139 9149 9159 9169 9179  
K G K G K S T A G F G M I F R H I A H G M P C A U  
AAGGGCAGGGCAGTCGACCCGGCTTCGGCATGATCTTCCGCCATATGCCACGGCATGCCCTGCGCCGTC  
9184 9194 9204 9214 9224 9234 9244 9254  
U Q F I K G A M A T G E R E L I E K H F G D U C Q  
GTGCAAGTCATCAAGGGTGCATGGCAGACCCGGCGAGCTTGATCGAGAAGCATTTCGGCGATGTCGCG  
9259 9269 9279 9289 9299 9309 9319 9329  
F Y T L G E G F T W E T Q D R A R D U A M A E K A  
TTCTACACGCTCGGCAGGGCTCACCTGGGAAACGCAAGGATCGCGCCCGCGATGTTGCGATGGCTGAAAGGCC  
9334 9344 9354 9364 9374 9384 9394 9404  
W E K A K E L I R D E R N S M U L L D E I N I A L  
TGGGAGAAGGGCGAGGAACGATCCGTGACGAGCGCAACTCGATGGTGCTCGACGAGATCACATTGCTCTG  
9409 9419 9429 9439 9449 9459 9469 9479  
R Y D Y I D U A E V U R F L K E E K P H M T H U U  
CGCTACGACTACATCGACGTGCCGAGTGGTGCCTCTGAGGGAGAAAGCCGACATGACGCGATGGTG  
9484 9494 9504 9514 9524 9534 9544 9554  
L T G R N A K E D L I E U A D L U T E M E L I K H  
CTCACCGGCCGCAACGCGAAGAGAGACCTGATCGAGTCGCCGATCTCGTCACTGAGATGGAGCTGATCAGCAT  
9559 9569 9579 9589 9599 9609 9619 9629  
P F R S G I K A Q Q G U E F \*  
CCGTTCCGTTCCGGCATCAAGGGCGAGCAGGGCGTGGAGTTCTGA  
9634 9644 9654 9664 9671

FIGURE 47.10

PROTEINE COBO			PREMIER RESIDU =	1	141/151
			DERNIER RESIDU =	214	
			NOMBRE	X NOMB	POIDS
1	PHE	F	9	4.21	1323.62
2	LEU	L	12	5.61	1357.01
3	ILE	I	14	6.54	1583.18
4	MET	M	11	5.14	1441.45
5	VAL	V	15	7.01	1486.03
6	SER	S	4	1.87	348.13
7	PRO	P	4	1.87	388.21
8	THR	T	13	6.07	1313.62
9	ALA	A	23	10.75	1633.85
10	TYR	Y	3	1.40	489.19
11	*	*	0	0.00	0.00
12	HIS	H	7	3.27	959.41
13	GLN	Q	5	2.34	640.29
14	ASN	N	4	1.87	456.17
15	LYS	K	21	9.81	2689.99
16	ASP	D	13	6.07	1495.35
17	GLU	E	23	10.75	2967.98
18	CYS	C	2	0.93	206.02
19	TRP	W	2	0.93	372.16
20	ARG	R	12	5.61	1873.21
21	GLY	G	17	7.94	969.36
22	-	-	0	0.00	0.00

RESIDUS = 214  
 Masse moleculaire (monoisotopique) = 24012.2500  
 Masse moleculaire (moyenne) = 24027.6973  
 INDEX DE POLARITE (%) = 47.66  
 POINT ISOELECTRIQUE = 6.94  
 DO 260 (mg/ml) = 0.443 DO 280 (mg/ml) = 0.612

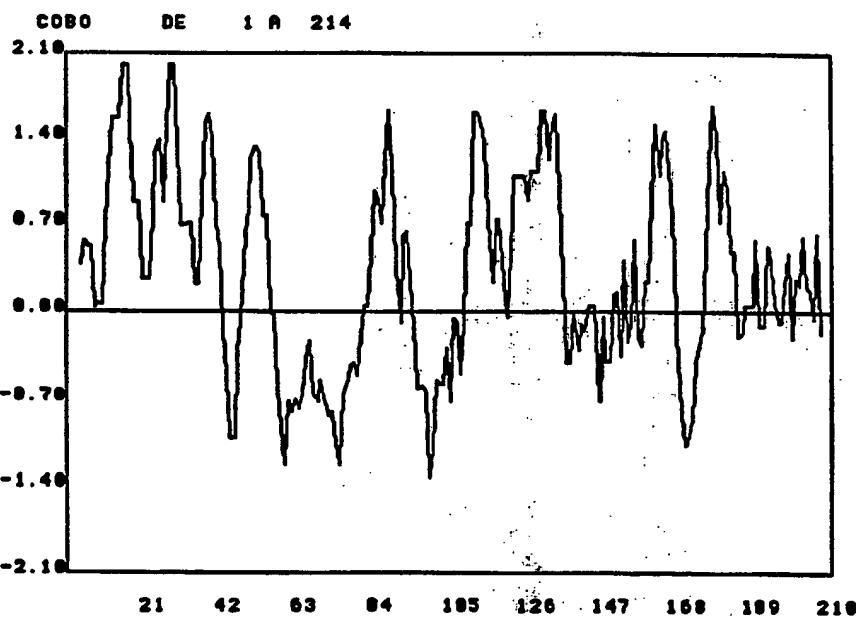


FIGURE 47.11

**A**

PL. 142 (15)

Séquence NH<sub>2</sub>-terminale de la SUMT de M. ivanovii

VVYLVGAGPGDPELITLKAVNVLK-ADVVL  
 923      946

Oligonucléotide sens 946 (27-mers)

P	G	D	P	E	L		
5' CGCGGAATTC	CCA	GGA	GAT	CCA	GAA	CT	3'
EcoRI	T	T	C	T	G		
	C	C	C				
	G	G	G				

Oligonucléotide sens 923 (27-mers)

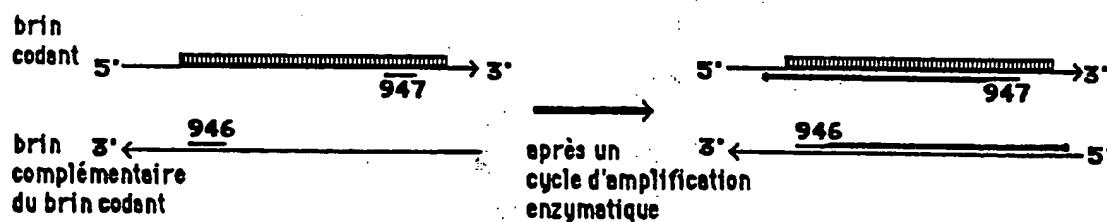
V	Y	L	V	G	A		
5' CGCGGAATTC	GTA	TAT	CTA	GTA	GGA	GC	3'
EcoRI	G	C	T	T	T		
				G			

Séquence NH<sub>2</sub>-terminale d'un fragment issu d'une digestion trypsique de la SUMT de M. ivanovii

IITGTLENIAGK  
 947

Oligonucléotide antisens 947 (25-mers)

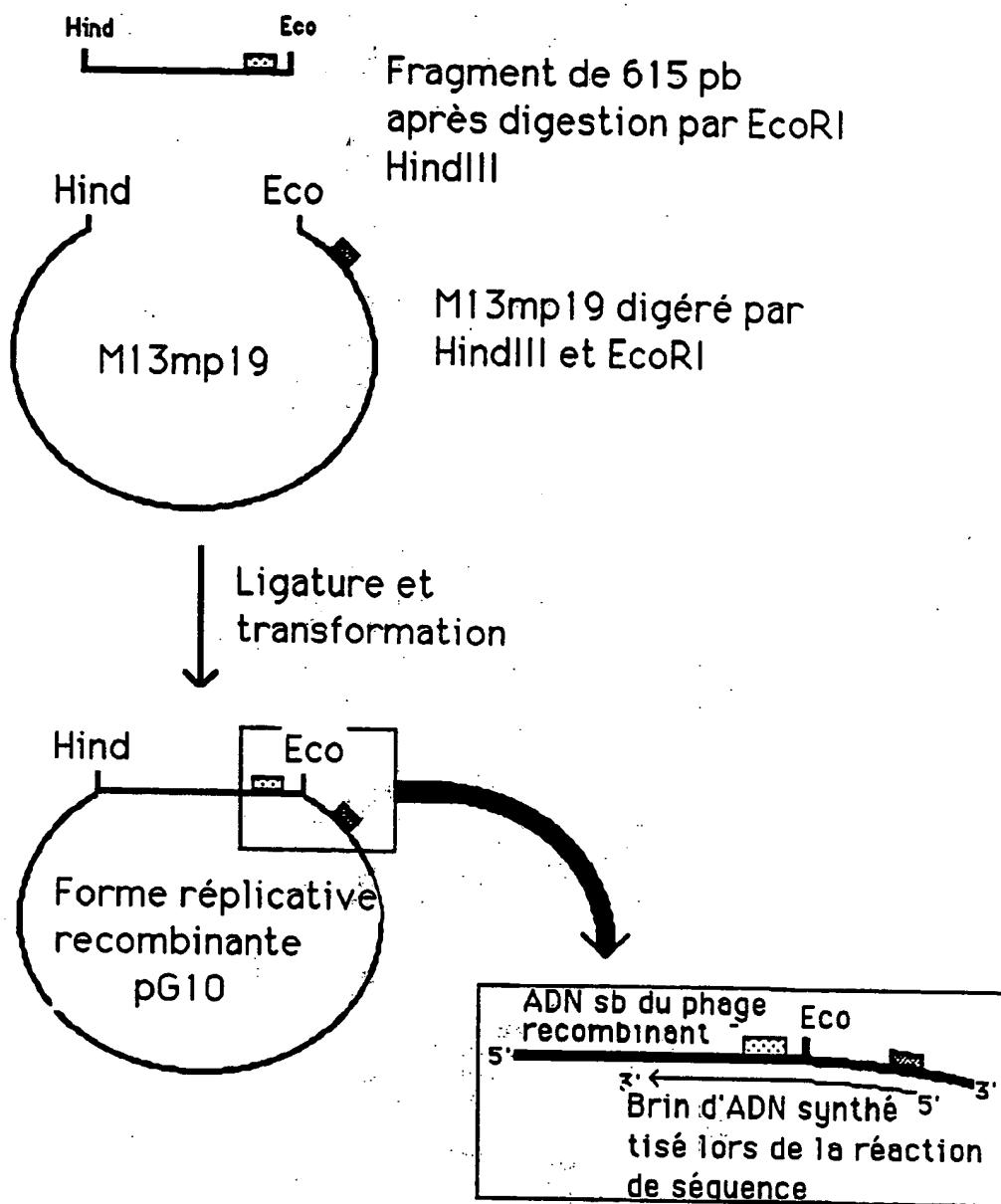
N	E	L	T	G		
5' CGCGAAGCTT	GTT	TTC	TAG	AGT	ACC	3'
HindIII	A	C	A	T	T	
				G	C	
					G	

**B**

■■■■■ gène de structure de la SUMT de M. ivanovii

FIGURE 48

143 / 151



- Site d'hybridation du primer -20 de l'ADN sb du phage M13mp19
- Séquence complémentaire de l'oligonucléotide sens 946

FIGURE 49

PL. 144 | (57)

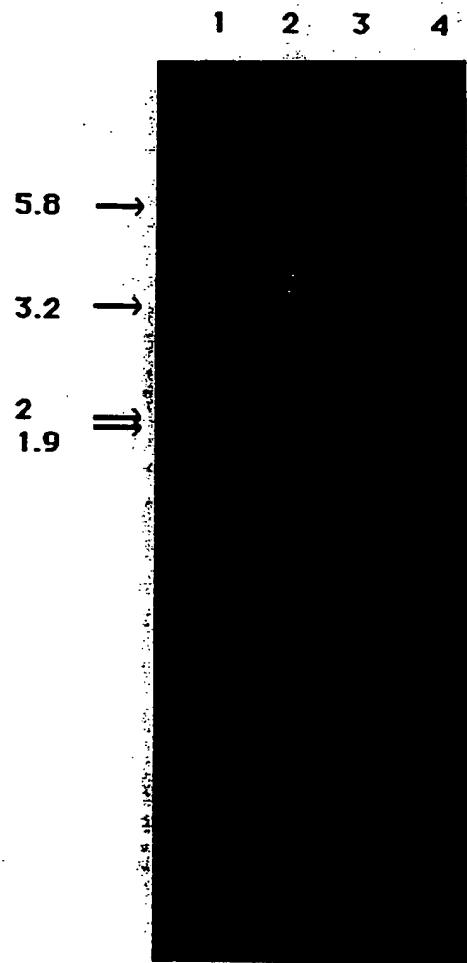


FIGURE 50

• PLT 145 / 151

10            20            30            40            50            60  
 CCATAATTCT TTTATAATT AAACGGTGAA CACATGGTAG TTATTTAGT AGGTGCGGGT  
 GGTATTAAGA AAATATTAAA TTGCACTT GTGACCATC AAATAATCA TCCACGCCA  
  
 70            80            90            100          110          120  
 CCAGGAGATC CCGAACCTAT CACTCTAAA GCTGTAAACG TGTTAAAAAA AGCGGATGTT  
 GGCTCTAG GGCTTGAATA GTGAGAGTT CGACATTTCG ACAATTTCG TCGCCTACAA  
  
 130          140          150          160          170          180  
 GTACTGTACG ACACACCTGC AAATGAAGAA ATTTAAAGT ATGCTGAAGG TGCAAAACTA  
 CATGACATGC TGTTCGACG TTACTTCCT TAAATTCGA TACGACTTCG ACGTTTGAT  
  
 190          200          210          220          230          240  
 ATATATGTCG GAAAACAAGC AGGACATCAT TACAAATCTC AAAATGAAAT CAATACTCTT  
 TATATACAGC CTTTGTTCG TCCTGTAGTA ATGTTAGAG TTTTACTTTA GTTATGAGAA  
  
 250          260          270          280          290          300  
 CTTGTGAAG AAGCAAAAGA AAATGATTAA GTAGTACGCC TTAAAGGTGG AGACCCCTTT  
 GAACAACCTC TTCGTTTCT TTACTAAAT CATCATGCCG AAATTCCACC TCTGGGAAA  
  
 310          320          330          340          350          360  
 GTATTTGGAA GAGGAGGCGA GGAAATTCTG GCCCTTGTAG AAGAAGGAAT TGATTTGAG  
 CATAAACCTT CTCCCTCCGCT CCTTTAACAC CGGGAACATC TTCTTCCTTA ACTAAAACTC  
  
 370          380          390          400          410          420  
 TTAGTTCCAG GGGTAACCTC TGCAATTGGA GTTCCAACAA CAATTGGCT TCCAGTTACT  
 AATCAAGGTC CCCATTGAAG ACGTTAACCT CAAGGTTGTT GTAAACCCGA AGGTCAATGA  
  
 430          440          450          460          470          480  
 CATAGAGGTG TTGCAACATC GTTACAGTT GTTACAGGTC ATGAAGACCC AACAAAATGC  
 GTATCTCCAC AACGTTGTAG CAATGTCAA CAATGTCCAG TACTTCTGGG TTGTTTACG  
  
 490          500          510          520          530          540  
 AAGAAACAGG TAGGATGGGA CTTTAAAGCA GATACTATTG TAATACCTTAT GGGTATTGGA  
 TTCTTGTCC ATCCTACCTT GAAATTTCGT CTATGATAAC ATTATGAATA CCCATAACCT  
  
 550          560          570          580          590          600  
 AATTTAGCTG AAAATACAGC AGAAATTATG AAACATAAAAG ATCCTGAAAC TCCAGTTGT  
 TTAATCGAC TTTTATGTCG TCTTTAACAC TTGTATTTC TAGGACTTGT AGGTCAAACA  
  
 610          620          630          640          650          660  
 GTAATTGAAA ATGGTACGAT GGAAGTCAA AGGATAATAA CGGGCACACT GGAAAATATA  
 CATTAACCTT TACCATGCTA CCTTCCAGTT TCCTATTATT GCGCGTGTGA CCTTTATAT  
  
 670          680          690          700          710          720  
 GCTGGAAAGG ATATTAACCC TCCTGCTTTA GTGGTATTGG AAATGTTGTC AATGTTTTA  
 CGACCTTCC TATAATTGAG AGGACGAAAT CACCATAACC TTACAAACAG TTACAAAAAT  
  
 730          740          750          760          770          780  
 AAGAAATGAA TCAAATCAGT GGCTGATCTA TAAAGAAGGC AATATCATGA ATGGATTAGA  
 TTCTTACTT AGTTTAGTCG CCGACTAGAT AATCTTCG TTATAGTACT TACCTAATCT  
  
 790          800          810          820          830          840  
 AGGAAAAAAA ATTGTTATAA CAAGACCTGC TGAAAGGGCT AAAGACTCAG TTGAATGGT  
 TCCATTTTT TAACAATATT GTTCTGGACG ACTTTCCGA TTCTGAGTC AACTTTACCA  
  
 850          860          870          880          890          900  
 AAAATCTTAT GGAGCAGTC CAATTGTAAC TCCTACATT GAACTCAAAG ATTCCAAGCC  
 TTTAGAATA CCTCGTCAAG GTTAAACATTG AGGATGTTA CTTGAGTTTC TAAGGTTCGG  
  
 910          920          930          940          950          960  
 AGAAGAAGTG ATAAAATTAT GTATATGAT AAATGAACT TGATTGGCT TATAT  
 TCTCTTCAC TATTTTAATA CATTACTA TTACTTGGG ACTAACCGGA ATATA

FIGURE 51

PL. 146 / 151

GENE CORA ET PROTEINE CORA (SUMT) DE METHANOBACTERIUM  
 IVANOVII  
 SEQUENCE DU FRAGMENT DE 955 PB DE 34 A 729

M	V	V	Y	L	V	G	A	G	P	G	D	P	E	L	I	T	L	K	A	V	N	V	L	K	
ATGGTAGTTTATTAGTAGGTGCGGGTCCAGGAGATCCCGAACTTATCACTCTCAAAGCTGTAAACGTGTAAAAA																									
34	44	54		64		74		84		94		104													
K	A	D	V	V	L	Y	D	K	P	A	N	E	I	L	K	Y	A	E	G	A	K	L	I		
AAAGCGGATGTGTACTGTACGACAAACCTGCAAATGAAGAAATTAAAGTATGCTGAAGGTGCAAAACTAATA																									
109	119	129		139		149		159		169		179													
Y	V	G	K	Q	P	G	H	H	Y	K	S	Q	N	E	I	N	T	L	L	V	E	E	A	K	
TATGTCGGAAAACAAGCAGGACATCATTACAAATCTCAAATGAAATCAATACTCTCTTGTTGAAGAACGAAAAA																									
184	194	204		214		224		234		244		254													
E	N	D	L	V	V	R	L	K	G	G	D	P	F	V	F	G	R	G	G	E	E	I	L	A	
GAAAATGATTAGTAGTACGCCCTAAAGGTGGAGACCCCCTTGTATTGGAGAGGGAGGGCGAGGAATTCTGGCC																									
259	269	279		289		299		309		319		329													
L	V	E	E	G	I	D	F	E	L	V	P	G	V	T	S	A	I	G	V	P	T	T	I	G	
CTTGTAGAAGAAGGAATTGATTGAGTTAGTTCCAGGGTAACCTCTGCAATTGGAGTTCCAACAAACAATTGGG																									
334	344	354		364		374		384		394		404													
L	P	V	T	H	R	G	V	A	T	S	F	T	V	V	T	G	H	E	D	P	T	K	C	K	
CTTCCAGTTACTCATAGAGGTGTTGCAACATCGTTTACAGTTGTTACAGGTATGAAGACCCACAAAATGCAAG																									
409	419	429		439		449		459		469		479													
K	Q	V	G	W	D	F	K	A	D	T	I	V	I	L	M	G	I	G	N	L	A	E	N	T	
AAACAGGTAGGGACTTTAAAGCAGATACTATTGTAATACTTATGGTATTGGAAATTAGCTGAAATACA																									
484	494	504		514		524		534		544		554													
A	E	I	M	K	H	K	D	P	E	T	P	V	C	V	I	E	N	G	T	M	E	G	Q	R	
GCAGAAATTATGAAACATAAAGATCCTGAAACTCCAGTTGTATTGAAATGGTACGTGGAAAGGTCAAAGG																									
559	569	579		589		599		609		619		629													
I	I	T	G	T	L	E	N	I	A	G	K	D	I	K	P	P	A	L	V	V	L	E	M	L	
ATAATAACGGGCACACTGGAAAATATAGCTGGAAAGGATATTAAACCTCCTGCCTTGTGGTATTGGAAATGTTG																									
634	644	654		664		674		684		694		704													
S	M	F	L	K	K	*																			
TCAATGTTTAAAGAAATGA																									
709	719	729																							

FIGURE 52

FEUILLE DE REMplacement

## PROTEINE CORA

144/151

			NOMBRE	% NOMB	POIDS	% POIDS
1	PHE	F	6	2.60	882.41	3.54
2	LEU	L	22	9.52	2487.85	9.99
3	ILE	I	17	7.36	1922.43	7.72
4	MET	M	6	2.60	786.24	3.16
5	VAL	V	26	11.26	2575.78	10.34
6	SER	S	4	1.73	348.13	1.40
7	PRO	P	12	5.19	1164.63	4.68
8	THR	T	16	6.93	1616.76	6.49
9	ALA	A	16	6.93	1136.59	4.56
10	TYR	Y	5	2.16	815.32	3.27
11	*	*	0	0.00	0.00	0.00
12	HIS	H	5	2.16	685.29	2.75
13	GLN	Q	4	1.73	512.23	2.06
14	ASN	N	9	3.90	1026.39	4.12
15	LYS	K	20	8.66	2561.90	10.29
16	ASP	D	11	4.76	1265.30	5.08
17	GLU	E	21	9.09	2709.89	10.88
18	CYS	C	2	0.87	206.02	0.83
19	TRP	W	1	0.43	186.08	0.75
20	ARG	R	4	1.73	624.40	2.51
21	GLY	G	24	10.39	1368.52	5.50
22	-	-	0	0.00	0.00	0.00

RESIDUS = 231

Masse moleculaire (monoisotopique) = 24900.1855

Masse moleculaire (moyenne) = 24915.9766

INDEX DE POLARITE (%) = 40.69

POINT ISOELECTRIQUE = 5.45

DO 260 (1mg/ml) = 0.304 DO 280 (1mg/ml) = 0.462

FIGURE 52.1

PL. 148/151

CORA DE PREMIER RESIDU=1  
DERNIER RESIDU=231

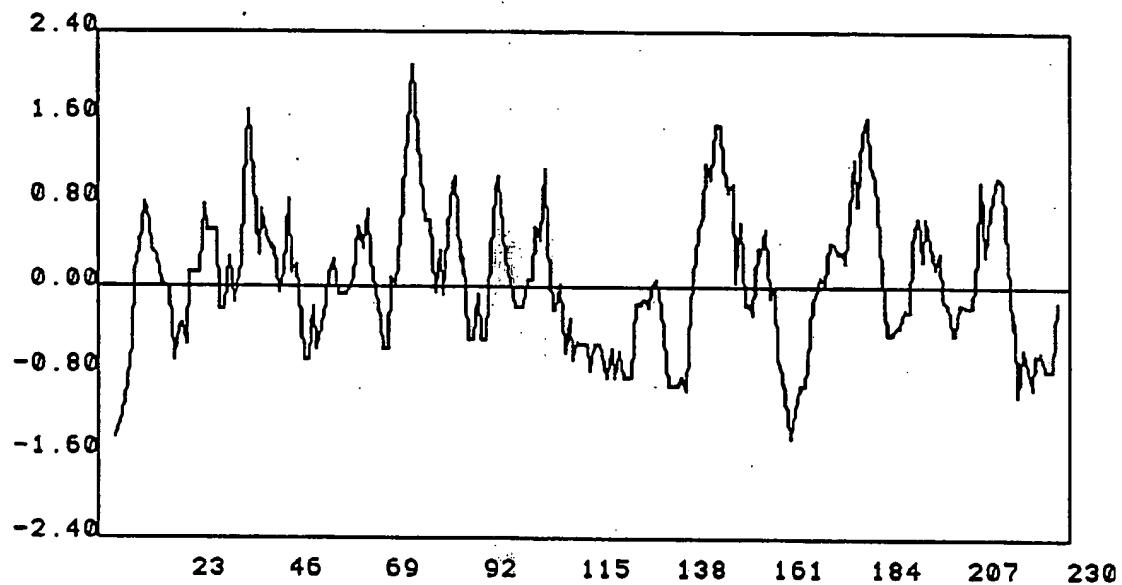


FIGURE 52.3

149/151

M. IVANOVII CORA                    3 A 227  
 P. DENITRIFICANS COBA    17 A 251

10	20	30	40	50	60	70
VYLVGAGPGDPPELITLKAVNVLKKADVVLYDKPANEELKYAE-GAKLIYVGKQAGHHYKSQNEINTLLV						
20	30	40	50	60	70	80
VWLGVAGPGDPGLLTLHAANALRQADVIVHDALVNEDCLKLARPGAVLEFAGKRGKPSPKQRDISLRLV						
80	90	100	110	120	130	140
EEAKENDLTVRLKGDPFVFGRGEEILALVEEGIDFELVPGVTSAIGVPTTIGLPVTHRGVATSFTVVT						
90	100	110	120	130	140	150
ELARAGNRVRLKGDPFVFGRGEEALTIVHQPFRIVPGITAGIGGLAYAGIPVTHREVNHAVTFLT						
150	160	170	180	190	200	200
GHEDPTKCKQVGWD-FKADT-IVIL-MGIGNLAENTAEIMKH-KDPETPVCVIENGTMEGQRIITGTL-						
160	170	180	190	200	210	220
GHDSSGLVPDRINWQGIASGSPVIVMYMAMKHIGAITANLIAGGRSPDEPVAFVCNAATPQQAVLETTLA						
210	220					
--E-NIAGKDIKPPALVVL-EMLSM						
RAEADVAAAGLEPPAIVVVGEVVR						
230	240	250				

POURCENTAGE D'HOMOLOGIE STRICTE : 40.4

FIGURE 53

150 / 151

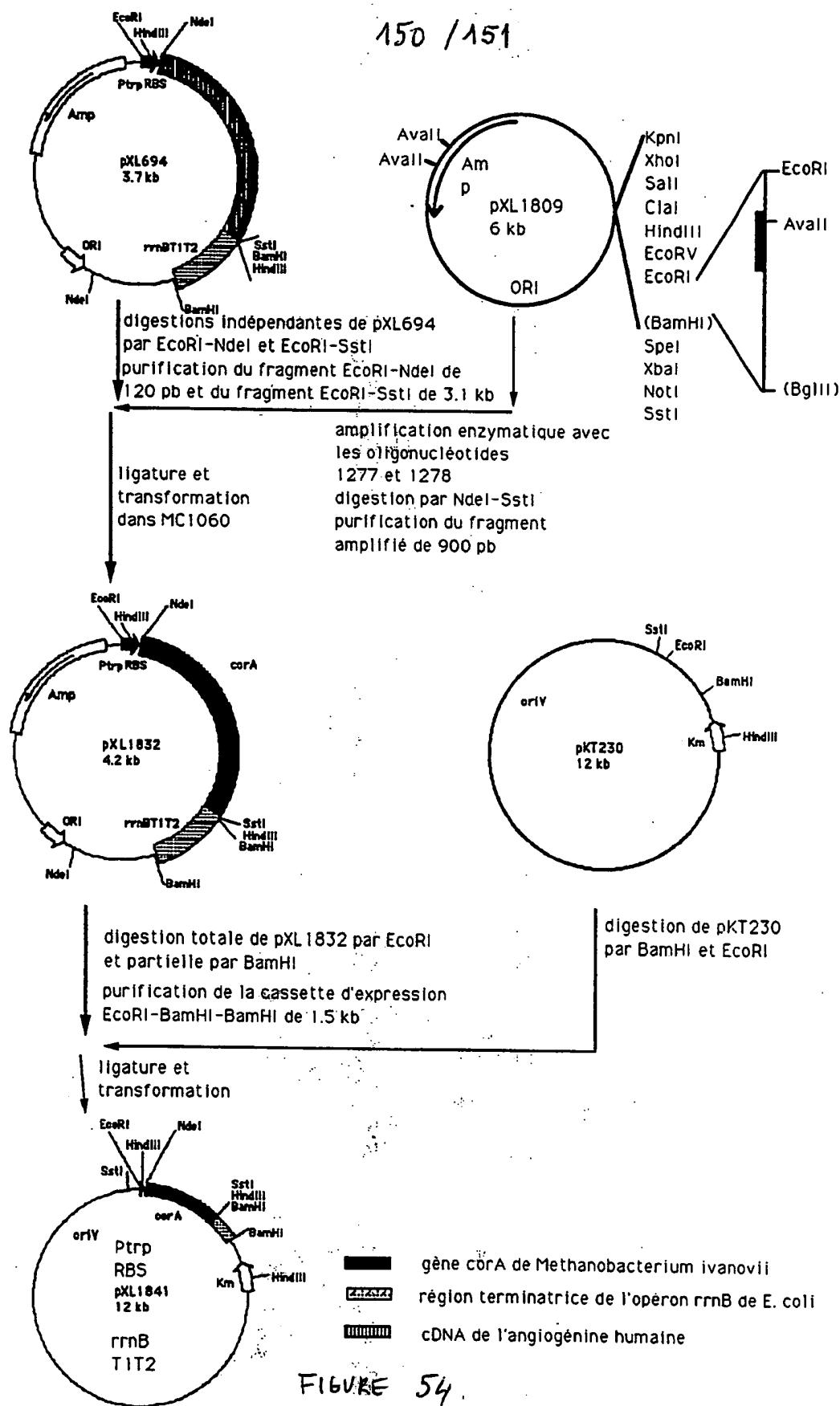


FIGURE 54

tableau: purification de la cobinamide kinase-cobinamide phosphate guanylyltransferase de *Pseudomonas denitrificans*.

Etape de purification	vol (ml)	protein (mg)	Cobinamide kinase		Cobinamide phosphate guanylyltransferase		ratio des act. spe. 2/1
			act spe 1 (U/mg de protein)	rendement (%)	facteur de purification	act. spe. 2 (U / mg de protein)	
Extrait brut <sup>a</sup>	4.5	120	16	-	-	214	13
Eluat Mono Q HR 10/10	9.0	8.98	188	88	12	-	-
Eluat Hydroxyapatite	2.0	4.55	325	77	20	3640	311
Eluat Phenyl-Superose	2.0	1.51	560	44	35	-	-
Eluat Mono Q HR 5/5	3.0	0.90	786	37	49	11282	14

<sup>a</sup>A partir de 1 g de cellules humides de SC510 pXL622 cultivées sur milieu PS4 (Cameron et al., 1989) sans cobalt.

TABLEAU A

# INTERNATIONAL SEARCH REPORT

International Application No PCT/FR 91/00054

## I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) \*

According to International Patent Classification (IPC) or to both National Classification and IPC  
 Int.Cl. 5 C 07 K 15/04, C 12 P 19/42, C 12 N 9/00, C 12 N 1/21,  
 // (C 12 P 19/42, C 12 R 1:38, 1:41, 1:01)

## II. FIELDS SEARCHED

Minimum Documentation Searched	
Classification System	Classification Symbols
Int.Cl. 5	C 12 N, C 12 P, C 07 K
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched	

## III. DOCUMENTS CONSIDERED TO BE RELEVANT

Category * 1	Citation of Document, * 2 with indication, where appropriate, of the relevant passages * 3	Relevant to Claim No. * 4
X	<p>Journal of Bacteriology, volume 171, No 1, January 1989, American Society for Microbiology, (Washington, DC, US)          B. Cameron et al.: "Cloning and analysis of genes involved in coenzyme B12 biosynthesis in <i>Pseudomonas denitrificans</i>", pages 547-557          cited in the application</p> <p>---</p>	1-42
A	<p>Journal of Bacteriology, volume 171, No 8 August 1989, American Society for Microbiology, (Washington, DC, US),          F. Blanche et al.: "Purification and characterization of S-adenosyl-L-methionine: Uroporphyrinogen III methyltransferase from <i>Pseudomonas denitrificans</i>", pages 4222-4231          cited in the application</p> <p>---</p>	./.

\* Special categories of cited documents: \* 10

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"S" document member of the same patent family

## IV. CERTIFICATION

Date of the Actual Completion of the International Search

21 May 1991 (21.05.91)

Date of Mailing of this International Search Report

17 July 1991 (17.07.91)

International Searching Authority

European Patent Office

Signature of Authorized Officer

## III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	Journal of Bacteriology, volume 169, No 7 July 1987, American Society for Microbiology, (Washington, DC, US) R.M. Jeter et al.: "Cobalamin (vitamin B12) biosynthetic genes of <i>Salmonella</i> <i>typhimurium</i> ", pages 3189-3198 cited in the application ----	
A	Journal of Bacteriology, volume 167, No 2 August 1986, American Society for Microbiology, (Washington, DC, US), R.N. Brey et al.: "Cloning of multiple genes involved with cobalamin (vitamin B12) biosynthesis in <i>bacillus megate-</i> <i>rium</i> ", pages 623-630 cited in the application ----	
A	WO, A, 8701391 (AMGEN) 12 March 1987 ----	
P,X	Journal of Bacteriology, volume 172, No 10 October 1990, American Society for Microbiology, (Washington, DC, US) J. Crouzet et al.: "Nucleotide sequen- ce of a <i>Pseudomonas denitrificans</i> 5,4- kilobase DNA fragment containing five cob genes and identification of struc- tural genes encoding S-adenosyl-L- methionine: Uroporphyrinogen III methyltransferase and cobyric acid a,c-diamide synthase", pages 5968- 5979 ---	1-31
P,X	Journal of Bacteriology, volume 172, No 10, October 1990, American Society for Microbiology, (Washington, DC, US), J. Crouzet et al.: "Genetic and sequen- ce analysis of an 8,7-kilobase <i>Pseudo-</i> <i>monas denitrificans</i> fragment carrying eight genes involved in transformation of precorrin-2 to cobyric acid", pages 5980-5990 -----	1-31

**ANNEX TO THE INTERNATIONAL SEARCH REPORT  
ON INTERNATIONAL PATENT APPLICATION NO.**

**FR 9100054**

**SA 44527**

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 03/07/91. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A- 8701391	12-03-87	AU-B-	599046	12-07-90
		AU-A-	6229786	24-03-87
		EP-A-	0236429	16-09-87
		JP-T-	1500081	19-01-89

# RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale N° PCT/FR 91/00054

## I. CLASSEMENT DE L'INVENTION (si plusieurs symboles de classification sont applicables, les indiquer tous) \*

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB C 12 N 15/52,  
 CIB<sup>5</sup>: C 07 K 15/04, C 12 P 19/42, C 12 N 9/00, C 12 N 1/21,  
 // (C 12 P 19/42, C 12 R 1:38, 1:41, 1:01)

## II. DOMAINES SUR LESQUELS LA RECHERCHE A PORTÉ

Documentation minimale consultée \*

Système de classification	Symboles de classification
CIB <sup>5</sup>	C 12 N, C 12 P, C 07 K

Documentation consultée autre que la documentation minimale dans la mesure où de tels documents font partie des domaines sur lesquels la recherche a porté \*

## III. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS \*\*

Catégorie *	Identification des documents cités, ** avec indication, si nécessaire, des passages pertinents **	N° des revendications visées **
X	Journal of Bacteriology, volume 171, no. 1, janvier 1989, American Society for Microbiology, (Washington, DC, US) B. Cameron et al.: "Cloning and analysis of genes involved in coenzyme B12 biosynthesis in <i>Pseudomonas denitrificans</i> ", pages 547-557 cité dans la demande	1-42
A	Journal of Bacteriology, volume 171, no. 8, août 1989, American Society for Microbiology, (Washington, DC, US), F. Blanche et al.: "Purification and characterization of S-adenosyl-L-methionine: Uroporphyrinogen III methyltransferase from <i>Pseudomonas denitrificans</i> ", pages 4222-4231 cité dans la demande	./.

\* Catégories spéciales de documents cités: \*\*

- « A » document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- « E » document antérieur, mais publié à la date de dépôt international ou après cette date
- « L » document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- « O » document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- « P » document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- « T » document ultérieur publié postérieurement à la date de dépôt international ou à la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- « X » document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive
- « Y » document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- « & » document qui fait partie de la même famille de brevets

## IV. CERTIFICATION

Date à laquelle la recherche internationale a été effectivement achevée  
 21 mai 1991

Date d'expédition du présent rapport de recherche internationale

12.07.91

Administration chargée de la recherche internationale  
 OFFICE EUROPEEN DES BREVETS

Signature du fonctionnaire autorisé  
 Natalie Weinberg

III. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		(SUITE DES RENSEIGNEMENTS INDICUÉS SUR LA DEUXIÈME FEUILLE)
Catégorie*	Identification des documents cités, avec indication, si nécessaire, des passages pertinents	N° des revendications visées
A	Journal of Bacteriology, volume 169, no. 7, juillet 1987, American Society for Microbiology, (Washington, DC, US) R.M. Jeter et al.: "Cobalamin (vitamin B12) biosynthetic genes of <i>Salmonella typhimurium</i> ", pages 3189-3198 cité dans la demande ---	
A	Journal of Bacteriology, volume 167, no. 2, août 1986, American Society for Microbiology, (Washington, DC, US), R.N. Brey et al.: "Cloning of multiple genes involved with cobalamin (vitamin B12) biosynthesis in <i>bacillus megaterium</i> ", pages 623-630 cité dans la demande ---	
A	WO, A, 8701391 (AMGEN) 12 mars 1987 ---	
P,X	Journal of Bacteriology, volume 172, no. 10, octobre 1990, American Society for Microbiology, (Washington, DC, US) J. Crouzet et al.: "Nucleotide sequence of a <i>Pseudomonas denitrificans</i> 5,4-kilobase DNA fragment containing five cob genes and identification of structural genes encoding S-adenosyl-L-methionine: Uroporphyrinogen III methyltransferase and cobyric acid a,c-diamide synthase", pages 5968-5979 ---	1-31
P,X	Journal of Bacteriology, volume 172, no. 10, octobre 1990, American Society for Microbiology, (Washington, DC, US), J. Crouzet et al.: "Genetic and sequence analysis of an 8,7-kilobase <i>Pseudomonas denitrificans</i> fragment carrying eight genes involved in transformation of precorrin-2 to cobyric acid", pages 5980-5990 -----	1-31

**ANNEX TO THE INTERNATIONAL SEARCH REPORT  
ON INTERNATIONAL PATENT APPLICATION NO.**

**FR 9100054  
SA 44527**

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 03/07/91. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A- 8701391	12-03-87	AU-B-	599046	12-07-90
		AU-A-	6229786	24-03-87
		EP-A-	0236429	16-09-87
		JP-T-	1500081	19-01-89

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**